

MSU Working Papers in Second Language

Studies Volume 3, Number 1

2011

Editors' Message

Dear Reader,

On behalf of the [editorial team](#), we invite you to read the third issue of the MSU Working Papers in Second Language Studies, an open access publication. "Open access" means that all content is distributed freely online and is available to be read and enjoyed by everyone.

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.

In this issue of the Working Papers we are proud to include three interviews conducted by students in the Second Language Studies program. Each interview is with a distinguished researcher in our field. Betsy Lavolette discusses computer-assisted language learning with Dr. Julie Sykes, Le Anne Spino welcomes Dr. Bill Van Patten to MSU, and Yeon Heo talks shop with Dr. Rod Ellis.

This issue also contains two research articles. The first, by Wen-Hsin Chen, attempts to show how native English speakers and Chinese ESL learners respond differently to compliments. The second, by Ayman Mohamed, is a mixed-methods investigation of incidental vocabulary learning in English conversation classes.

This issue includes one book review: Yunson Shin reviews *Talk Time* student book 2: *Everyday English Conversation*, by Susan Stempleski.

Because the Working Papers is intended to show works in progress, we accepted two proposals this year. It is our hopes that after reading these proposals that you will send any helpful comments to the authors. Magdaleen Corne Lotter proposes a qualitative study of the role of reading and writing skills in the development of oral proficiency of young learners of English in Taiwan. Next, Hyojung Lim and Young-Shin Kwon show the amount of effort that goes into making and testing a test.

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, Russ Werner, who provided tech support for the Working Papers website, and SOSLAP, who for put out multiple call for papers and volunteers. Without the help of these tireless volunteers, the Working Papers would not be possible.

All of the following volunteers are MSU students, graduates, and alumni.

Copy editors:

- Laura Ballard
- Frances Lamielle
- Ayman Mohamed
- Ji-Hyun Park
- Le Anne Spino
- Jamie Thomas
- Cristen Vernon

Section editors:

- Wen-Hsin (Kelly) Chen
- Solène Inceoglu
- Hyo Jung Lim
- Le Anne Spino

We hope you enjoy reading the 2012 edition of the MSU Working Papers in Second Language Studies.

Elizabeth (Betsy) Lavolette
Scott Sterling
Co-Editors

Interview with Rod Ellis

Interviewed by Yeon Heo
PhD student, Second Language Studies Program
Michigan State University
heoyeon@msu.edu

Professor Rod Ellis is the deputy head of the Department of Applied Language Studies and Linguistics at the University of Auckland, New Zealand. He came to the U.S. as a Professor in the University's Distinguished Professor visitor program. He stayed in the U.S. for five weeks. He is planning to visit the U.S. again in April 2013 for a week. At Michigan State University, he taught a short course, "Planning and Task-Based Performance," and gave the PhD students insights into how to do research related to task-based learning/teaching.

His research interests include: second language acquisition, individual learner differences, form-focused instruction, teacher education, course design, and methodology of language teaching. He was kind enough to do the interview for the Second Language Studies Working Papers.

Could you tell me how you got involved in second language studies?

Well, I suppose that's quite a long story. There were really two influences that motivated me to get involved in second language studies. The first was that, like many second language acquisition researchers, I started off as a language teacher. I was a language teacher in Spain for a short time, and then I was a language teacher in a secondary school in Zambia in Africa. One of the things that I became increasingly aware of was the gap between teaching and learning; teachers tend to make certain assumptions that if you teach something well, learners will learn it. It became quite clear to me that very often no matter how much effort you put into trying to teach learners a particular grammatical structure, there was no guarantee that they would be able to use it correctly in their communicative speech or in their writing. So that got me interested in why there was this gap between teaching and

learning and how one could minimize the gap. I realized that this would involve investigating language learning. Of course, this was back in the early 1970's and there wasn't very much published on second language acquisition at that particular time. In fact, it still is a very new subject. So that was one of the influences.

And the second major influence occurred when I left Africa in 1970 and went back to the United Kingdom. I decided to do a Master in Education and worked with someone called Gordon Wells. Gordon Wells at that time was working on a child language acquisition project. Gordon introduced me to the exciting and interesting work that was going on in first language acquisition research. At that time, it became quite clear to me that a lot of the things people were doing in that line of inquiry were also going to be very relevant to inquiring about second language learning. So those were two huge inputs. One from my experience as a language teacher and the other from my

experience as a researcher in a master's program with Gordon Wells.

Would you briefly introduce your research interests?

I guess over the last few years I've had a number of research interests in the field of second language acquisition and its application to language pedagogy. One of the areas is how teachers focus on form in communicative language classrooms. Together with Shawn Loewen and Helen Basturkmen, I conducted a project in 1999 and 2000 where we investigated what we called "form-focused episodes" as those occurred in communicative language lessons in a private language school in Auckland, New Zealand. That eventually led to Shawn's PhD thesis. He took our research further by investigating to what extent form-focused episodes actually facilitate language learning.

Another area that I have had a prevailing interest in is corrective feedback—both oral and written corrective feedback. I've been involved in studies that have investigated the effects of different types of corrective feedback on both students' oral production and their writing. The third area is ways of measuring implicit and explicit knowledge. Together with a number of other people including Shawn Loewen, I conducted a study where we attempted to develop tests that would provide relatively separate measures of those two types of knowledge. The results were eventually published in a series of articles and a book. In fact, there's a copy of the book just outside this room here called *Implicit and Explicit Knowledge in Language Proficiency, Testing, and Teaching* (Ellis, Loewen, Elder, & Erlam, 2009). So those are the main areas that

I've been working on and they continue to be so.

Could you tell me what you're working on currently?

I do have a research project we've been working on fairly recently, which again is a corrective feedback study. What we wanted to do was to test some of the claims of socio-cultural theory as opposed to cognitive interactionist models of language learning, because they make somewhat different predictions about the kind of corrective feedback that is likely to be most effective in promoting learning. Sociocultural theory argues for a scaffolded approach where the teacher moves from relatively implicit types of correction to more explicit types of correction, finding the optimal type of correction for eliciting the correction from the individual student. In contrast, cognitive interactionist theories are more concerned with trying to identify the particular type of corrective feedback that is likely to work for all students. Indeed, researchers in this tradition have reported that explicit feedback works better than more implicit types of feedback. So we carried out a study where some students were subjected to the scaffolded approach to doing corrective feedback while other students were given direct, explicit corrective feedback. We were interested to see whether in fact there were any differences in learning outcomes. In fact, we found none. There was no clear evidence that explicit or scaffolded feedback was better. This does raise a question about the claims made about scaffolded feedback because scaffolded feedback is very time-consuming. Teachers have got to weigh the various strategies in an attempt to find the least explicit one to elicit a correction from a

learner. As a result, corrected episodes tend to be long. In contrast, when you're providing explicit feedback, the episodes are much shorter. So, you might want to argue that explicit feedback is more efficient than scaffolded feedback.

I have a lot of PhD students who are working on a variety of areas. One PhD student just completed a thesis looking at oral corrected feedback, specifically comparing two types of implicit feedback: recasts and requests for clarification. Interestingly there have been no studies that have actually compared an input-based implicit strategy such as recasts with an output prompting implicit type of corrective feedback such as requests for clarification. Both are implicit, but they differ in terms of whether they provide learners with the correct form or whether they elicit the correct form from the learner. This study was carried out in high school French classrooms in Auckland. Interestingly, what she found was somewhat different from what Lyster found in his research. She found recasts were considerably more effective than the clarification requests. So this study in a way challenges Lyster's claims that output prompting corrective feedback is more effective than recasts.

What constitutes a good researcher, say, a good PhD student?

I think that PhD students need certain skills and certain qualities in order to be really effective researchers. One of the things that I've noticed amongst my PhD students is a difference in how they look for and handle information. This concerns the well-known distinction between divergent thinkers and convergent thinkers. Divergent thinkers are often quite creative because they tend to see things

in ways other than those that are well-trod and well-established. On the other hand, they often find it quite difficult to develop a well-structured, coherent proposal. Students who are more convergent typically are much better at defining their research questions carefully, working out how to operationalize them, working out what kind of data they need, and how they are going to analyze the data, etc. Convergent thinkers often tend to elect for a more experimental quantitative approach whereas more divergent thinkers tend to opt for qualitative research. They both have strengths and weaknesses. I think one thing that PhD students have got to decide for themselves is what kind of person are they—how do they do their thinking. Do they tend to think holistically and in divergent ways or do they tend to be convergent in their thinking?

There can be cultural differences as well. For example, my experience of a lot of Asian students is that they tend to have a preference for quantitative, experimental, or correlational types of research rather than research that involves collecting data from a variety of different sources, looking for themes, and trying to analyze the themes, etc. But some people from other cultures tend to prefer a more qualitative approach. So I think students need to think very carefully about how they see the world, how they think, and then pick a research style that is going to suit them.

Probably the other quality PhD students need above everything is persistence. Not giving up. Not allowing themselves to get distracted by other things—keeping focused on their research, having a really clear schedule, a really good plan and trying to keep to the schedule and the plan. If you don't do that, you end up being a PhD student

for six–seven years and start to wonder about if you are ever going to complete. I never like my PhD students to take longer than four years because after that, I think there is a likelihood that they will not complete. They are wasting my time as well!

Another thing is good PhD students in my experience are not necessarily the ones who always do what I think they should do. I never require students to do what I think they should do. Very often they come up with something that is actually as good, if not better, than I have thought of. But I do expect them to listen carefully, to pay close attention to what I suggest. Perhaps, another characteristic of a good PhD student is that when they come to visit me to discuss their research they have worked out exactly what it is they wanted to discuss. I don't like students who just expect me to tell them what to do. They need to have very specific questions and very specific problems that they want addressed.

Do you think the qualities of a good language teacher and the qualities of a good researcher can be combined?

I suppose your question is addressing to what extent a teacher should engage in research and also conversely to what extent a researcher should engage in teaching. There is quite a big literature that encourages teachers to do research of various kinds, at least action research, but there's not much in the literature that actually talks about whether researchers should do language teaching. Maybe that's the topic that is worth investigating! I have to admit that I haven't done any language teaching for a number of years now. Although I do feel that the early part of my life, where I was a language teacher for many years,

has been foundational. I continually draw on that experience in terms of what I think as a researcher, etc.

Teachers becoming researchers? I think we probably have to recognize that this is an ideal. I wonder, for example, if you were to give out a questionnaire to the teachers in the ELC program here, and ask them their views about whether they should do research, what they would say. I suspect you will find the vast majority is not doing any research. Perhaps teachers don't typically do research. I think that there are two main reasons: time and motivation. Teachers are busy and research takes time. If you try to do research, you're going to make your life even busier. And motivation, I think a lot of teachers are perhaps skeptical as to whether research is actually going to help in their teaching. They may think there are other things they can do that will help them more to become a better teacher than doing research.

I think it's also important to ask about what kind of research teachers might do. Dick Allwright proposed something called 'exploratory practice' (Allwright, 2003). He argues that teacher research should not really be focused on problems or research questions, but rather should look at sort of what he calls 'puzzles'—things that teachers are not clear about, or why is something happening in their classrooms, or why something is not happening in the way in which they want it to happen. Allwright's idea is that exploratory practice is something that teachers and learners do collaboratively. They have to be jointly involved in trying to understand a puzzle. I think what motivates him is the idea that an understanding of what is going on in a classroom is actually much more important than collecting data and

answering specific research questions. Allwright has developed a series of principles to guide exploratory practice and has published these in a number of different papers.

I have argued that one way teachers can do ‘research’ is by focusing on the instructional materials they use and how they implement them. It seems to me that teachers make certain assumptions that if they use a particular type of activity, it will contribute to learning in a certain way, or it will induce a certain type of interaction, a certain type of language learning behavior in the classroom. So a very practical type of research that teachers can do is to sometimes carry out what I call micro-evaluations of specific teaching activities. I’ve tended to focus this on “tasks” because of my interest in task-based language teaching. I get my postgraduate students to design a task that they could use in a particular teaching context and to plan an evaluation of it. They then have to teach the task and carry out the evaluation, and write up a report of it. It’s time-consuming but my students report that they learn a lot by carrying out such evaluations.

Do you have any hobbies?

I spend a lot of time working! But I also do enjoy cooking. I do nearly all the cooking in my family. My partner does the washing up and I do the cooking! Cooking is very relaxing and also kind of creative. You have to think about how you can put together a tasty meal with whatever you happen to have in your fridge. Maybe when I finally retire—if ever I do—one of the things I’ll do is take

a cookery course so that I can improve myself as a cook.

Before we end the interview, do you have anything that you would like to add?

It’s been a very pleasant time staying at Michigan State University. I’ve enjoyed teaching my little course. I’ve enjoyed meeting some people, being able to participate in a research project, collecting some data here. One of the really nice things about coming to live somewhere different for a period of time is that your lifestyle changes. My lifestyle here is built around the fact that I have no television, I have no car, I have no telephone. I do have the internet, so I am not totally unable to communicate with people. But believe me, when you remove those three things from your life, your life changes! I walk everywhere, which is very good and healthy. I guess when I go back to Auckland, I will be getting in my car and driving to work and driving to go shopping, etc. So I’ve enjoyed coming here because for a while I’ve been able to change my lifestyle. It’s so easy to get stuck in one’s lifestyle and it’s really good to change it!

References

- Allwright, D. (2003). Exploratory practice: Rethinking practitioner research in language teaching. *Language Teaching Research*, 7, 113–142.
- Ellis, R., Loewen, S., Elder, S., & Erlam, R. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Clevedon, England: Multilingual Matters.

Interview With Dr. Bill VanPatten

Interviewed by Le Anne Spino
PhD student, Second Language Studies Program
Michigan State University
spinole@msu.edu

Bill VanPatten is a Professor of Spanish & Second Language Studies at Michigan State University. His research interests include parsing and processing, input processing, the interface between morphology and syntax, and instructed SLA. He is the author or co-author of 8 eight books and almost 100 articles. He has also authored Spanish and French language textbooks. This interview was conducted on March 19, 2012. For more information about Dr. VanPatten, please visit his website: <https://sites.google.com/site/bvpsla/>.

How did you first become interested in second language acquisition?

I became interested in second language acquisition during two distinct points in my life. I was raised in a bicultural bilingual family so I always had interest in people who knew more than one language. I had 55 cousins on my mother's side and in that group, that cohort, I think there were only three or four of us who spoke Spanish. We're all English dominant because we live in the United States but the rest of my cousins were so English dominant that they basically spoke no Spanish. Some understood to greater or lesser degrees but they didn't speak. So only three or four of us actually had fluency with the language. That interested me. I wanted to know why that was so... And then I got interested in more second language things when a friend of mine was working on a dissertation and asked me to teach in an experiment that she was running for her dissertation. That's where I first started learning about second language acquisition. It coincided with my first course on child language acquisition so it all came together at the same time.

Could you briefly describe the projects you're working on?

I just completed a major project on aptitude, actually grammatical sensitivity, and the processing part of processing instruction in four different languages. We're going to follow that up in two languages now with working memory as an individual difference. In another study we're working on a follow-up to a study that's coming out in June. It's a study on the relationship between syntax and morphology with verb-movement and person/number inflections on verbs. Our first study only focused on sensitivity to grammatical violations and we're following it up with a cross-sectional study where we're including a production measure because we want to see if there's any relationship between sensitivity to violations and productive ability with these things that we looked at in our first study. Then there's a study that a former student and I are wrapping up on Japanese as a second language on parametric variation with head directionality. We're both from the old school of parameters and we believe that some of these old parameters that have fallen to the wayside are still actually useable for talking about language acquisition. This

next year, another group of people and I have two studies in the works that are about anaphoric reference and antecedent choice for null and overt subject pronouns in Spanish. So I'll stop there. I have other ones but those are the main ones.

Over the years you have most certainly published a great deal. Which of your publications do you believe has had the greatest impact and why?

I can tell you something that a student of mine actually pointed out to me a month ago. If you go to the Studies in Second Language Acquisition (SSLA) website, my 1990 SSLA article and my 1993 SSLA article with Teresa Cadierno are two of most cited articles in SSLA ever. So I could probably say that those have had the most impact. One deals with the focus on content and form at the same time, that's the 1990 one and the 1993 one launched a whole agenda on processing instruction. I think the processing instruction one has had a lot of impact because it turned people on their heads when it came to thinking about the nature of instructed SLA. In fact, it still turns some people on their heads. It creates quite a stir because no matter what lip service people give to input, there's still a lot of people out there who just don't want to believe that input is the way you get language in your head.

How do you believe the field of SLA has changed over the years?

That's a tough one and I'm going to get politically incorrect, I'm sure [laughter]. I think it's changed in that it's become so multifaceted and so diffused that we no longer have a common research agenda. People have

lost sight of the history of SLA and how we got started and what the fundamental questions are. You see a proliferation of theories now and some of this proliferation of theories is because of people looking at different things but thinking they're looking at the same things. So they fight and they argue about it and it's like arguing over... Let's say you're baking an apple pie and I'm baking a cherry pie. Even though we chose to bake different kinds of pies, we still argue about what's the best method for baking a pie. Obviously you don't bake an apple pie the same way you bake a cherry pie, but we still argue about it. And that's kind of what's happening in SLA. There are apples and cherries going on.

So what would you say are the challenges the field of SLA is facing today?

I think it's facing that challenge, the challenge of not bifurcation but multifurcation. It's so splintered that there's just not a lot of common ground anymore. Also, it's always faced a political challenge because SLA is a field, at least in the United States, that has been dependent a lot on literature departments and traditional language departments. I don't think that's the healthiest thing for SLA. So aside from its research agenda and theoretical orientation—the multifacetedness I was talking about before—I think it still faces a political challenge because SLA is always the thing that gets short-shrifted in colleges of arts and sciences.

As a final question, what advice do you have for second language researchers, especially those who wish to bridge the gap between theory and practice?

My advice is not to be in such a hurry to bridge the gap between theory and practice. I think that there's a myth in the field that somehow everything we do has to be related to practice or everything can be translated into practice and that's just not the case. I recommend you read a 1985 article published by Pasty Lightbown in *Applied Linguistics* called "Great expectations: Second-language acquisition research and classroom teaching" about the relationship between SLA theory research and practice. It's as relevant in 2012 as it was in 1985. I think that people are so big on the practice part they lose sight of the bigger picture, that there's so much that we don't know about SLA. Constantly trying to do practice makes us jump the gun. Look at my work on processing instruction. A lot of people do one or two studies and they're done. We've been doggedly working on processing instruction for years now in all kinds of ways, looking at different nooks and crannies and uncovering the variables and changing the testing method because I'm just not content with saying, "Here we've got some results therefore here's my opinion about how things happen." You need to constantly replicate, rethink your variables, go back and examine things and just not be so quick to jump on the practice wagon. The fact that we have so many theories of SLA should be a clue that you can't jump on the practice wagon right away. You've got to just do the groundwork and we're still doing the groundwork. How old is the field? If you count S. P.

Corder's 1967's publication "The significance of learners' errors" followed by Larry Selinker's 1972 "Interlanguage" paper, then about 45 or 40 years—that's not a very long time for a field to be in existence. So I think that we should just be patient. Just be patient.

Author's note: I would like to thank Bill VanPatten for participating in this interview. His vibrant personality and vast knowledge of SLA makes every conversation with him a pleasure. I believe this interview gives a glimpse of the prolific amount of high-quality work Dr. VanPatten has published throughout the years and also provides valuable insight into the past and present of SLA.

References

- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5(4), 161–170.
- Lightbown, P. (1985). Great expectations: Second-language acquisition research and classroom teaching. *Applied Linguistics*, 6(2), 173–189.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(3), 209–232.
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12(3), 287–301.
- VanPatten, B. & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15(2), 225–243.

Interview With Dr. Julie Sykes

Interviewed by Elizabeth (Betsy) Lavolette
PhD student, Second Language Studies Program
Michigan State University
betsy@msu.edu

Dr. Julie Sykes is Assistant Professor of Spanish and Hispanic Linguistics at the University of New Mexico. She delivered a keynote address at the Second Language Studies Symposium titled Digital game-mediated language learning: From ideas to implementation. This interview took place on 2/24/2012. Her website can be found at <http://www.unm.edu/~spanport/faculty/sykes/>.

Would you talk a little bit about your language learning and teaching background?

I started learning Spanish when I was in high school, so I traveled abroad. My parents were really awesome, and they let me go by myself or with programs. So I lived in Spain and Costa Rica, and that's how I started learning Spanish. I majored in Spanish and International Studies in my undergrad. So I knew I always wanted to do intercultural communication, I just didn't know how to do that. So the plan was to take the foreign service exam and study, and so I had a professor who said, why don't you just go to grad school and keep up your Spanish while you're studying for the exam, and I decided, why not. It turns out that I just fell in love with linguistics and pragmatics specifically, and intercultural communication. So, the rest is history. I did my MA and PhD in applied linguistics and SLA, and now, I work as assistant professor of Spanish and Portuguese, mostly SLA, at University of New Mexico. I also supervise 35 graduate students in the first four semesters of Spanish as the coordinator

of the language program. So, I get to do lots of work with training, methodology, lower-division courses, curriculum, and that kind of thing, which I actually really like as well. I research different acquisition pragmatics and intercultural communications.

Do you currently teach Spanish?

I don't. I just teach graduate courses in methodology. So I teach the typical sort of training course for TAs, and then I also teach technology, and I teach pragmatic acquisition, graduate courses mostly. I teach one undergraduate course in the summer, Hispanic linguistics mostly. I also supervise all those language courses, like I said.

I've noticed that you work a lot with a professor at the University of Arizona. How did that collaboration start?

Jon Reinhardt is the professor that I work with down there, and he and I were graduate students at the same time, went to the same conferences, and so we just built a network of good colleagues. We've both worked with Steve Thorne. Steve was Jon's dissertation director, and then Steve worked with me as a

favor to me, which was great. So, this natural collaboration developed from being interested in the same work, but also enjoying working together. So, I guess we both feel like cross-institutional collaboration is really important and really fun for both of us. You know, two minds are always better than one.

How did you get the idea for Mentira [a virtual reality game for teaching Spanish pragmatics that is played using iPod touches], where you play out in the world, and the game is part of the world?

I work with another professor at UNM, Chris Holden, who got his PhD at the University of Wisconsin-Madison in the Games, Learning, and Society Program. He has actually been building AR (augmented reality) games for a long time, mostly for middle school students in sciences and ecology. So, he had a lot of experience in games. He and I actually met at a games conference, Games+Learning+Society, the summer before we both started at UNM, so it was just coincidence that we both ended up at the same place. So, really, I started working on place-based games because of him. It was circumstance, and I got really excited about place-based games for language learning. Being in Albuquerque, place-based games for Spanish are relevant because we have many Spanish-speaking communities, and there's a very long history of Spanish and Hispanic culture in lots of different things in New Mexico, so it's kind of a natural place to do it.

Would you say that games for language learning is your main research interest?

It's one of my two primary areas. I'm really interested in intercultural

pragmatics. Even with my master's thesis, I started working in pragmatic acquisition. Digital games are a by-product of that and, in a sense, have become primary just because of their nature and newness in the field. They give us lots of fresh areas to look at and things we don't know. I've combined these two interests, and that's what I'm most interested in. How do we utilize innovative technologies, not just because they're innovative but rather, to overcome some of these challenges we've been facing? In the last three or four years, we've really witnessed this shift in the way communication is mediated, the way mobile phones are used, the way digital gaming interfaces with real life. Language, of course, is a part of that, so just by the nature of where the world is going, that's what I've been interested in. Pragmatics is always a part of any of that.

Is Croquelandia [a 3D simulation game for teaching Spanish pragmatics] finished at this point?

In terms of development, yes, definitely. We're not going to move forward in developing Croquelandia. I still have a huge amount of data that needs to be analyzed. I've published four or five articles on different aspects of that data. It was a first time attempt, really, at anything for language learning. We've seen Zon out of Michigan State, actually, but we haven't seen a lot of data on it, so it would be nice to get some data if we could. But because we hadn't seen a lot, we didn't really know what we were doing when we developed Croquelandia. I think we've learned much of what we can learn from that experience, and now it's time to build and start over. We'll do it very differently the second time around. I'm looking forward to doing that. Right now,

I'm tied up with place-based games, but it doesn't mean it's at the exclusion of the other kind. It's just a matter of time, resources, and platforms. Game design and game publication is actually quite intensive. It's like designing an intervention, but much more complicated. So we actually consider them publications in their own right.

That's interesting that you say that you consider those to be publications. Do tenure committees also view them that way?

Yes and no. Definitely not the same way they would a scholarly publication. But very much like creative writing, the same way poetry might count or a creative novel. That's generally the consensus on where those might fit. The language is sort of a report on where digital scholarship in general falls, and that's what I think we're trying to figure out.

So, when you talk about game design, what exactly does that mean? Does that mean that you're writing the code? Or does that mean that you're designing how it's going to work?

Game design happens in a lot of ways. I'm not a programmer, really. I do a little bit, dabbling here and there, but not really. I definitely couldn't do a 3D game, for example. So, mostly, the design I do is around storyboarding, dialog, characters, narrative, and feedback mechanisms, and everything related to that. It's very complicated. There's a whole process of game design and what things you need to consider. Jesse Schell at Carnegie Mellon has a great book on game design that inspires our work. That's just one example. There are quite a few out there. It's a matter of

starting from the bottom up. All the text has to be built and all the images. Aris, which is the game platform we design in, is nice. It's out of the University of Wisconsin, it's open source, it's free to use.

That's the platform you used to make Mentira?

Yes, that's the one we used to make Mentira and this new game we're working on. Chris Holden is on that design team as well for the team for Aris, so they built things into Aris to give us capabilities that we wouldn't have otherwise had. So that's a part of a design phase as well. We tell them, no, we need to be able to do this, for example. Right now, we're working on getting audio, something that's really important for languages. Can we get audio to work as well as text?

If you're not the one who's coding these games, there must be quite a bit of money needed to hire people to do it.

Right. Croquelandia was really nice in that there was a team of programmers looking for some content, and so we each brought something to the table. It wasn't free, of course, but it was part of the university's program time to work with these experimental technologies. In the case of Aris, Aris is free and open source, and you don't need any programming at all. So the coding money, of course, goes into Aris, but not into Mentira itself. We work with them very closely to help create exemplars that can be used to gain funding, and Aris has gained quite a bit of traction. It works very much like a normal open source project, off grant money, things like that. Basically, Mentira was built with \$10,000 of internal grant money from UNM. Graduate student time and

devices is what that paid for; that was it. It's been nice. We've been finding the work-arounds. We call it ecological design. What can we build for the least amount of money, but still make it compelling and interesting?

The games that are being developed in Aris are all over the world and have really shot up in number in the last year. So, and there's funding coming in from the Library of Congress. It's a very well supported platform, which is why we're interested in working with them. Again, we work really closely with them to hopefully ensure its continuance and its success. They've really stayed true to that open-source spirit thus far, which is exciting. You can download Aris for free on your iPhone right now. You can search for Mentira, and you can play Mentira. I can send you the codes and you can play without being in the place. We'd prefer not to publish the codes, mostly because we have students still playing. We're happy to share, but we don't want them to just Google the codes.

I think that one of the exciting things about CALL and about gaming is that it's always changing, and there's always something new to learn. But I've also heard people argue, how could you possibly research that? As soon as you publish something, it's on to the next thing. It's already old. So, what would be your response to that?

My response is two-fold. One, it's not about the tool. It's about the behaviors. Game-based behaviors have been around for a long time. We see them in more complicated, visually rendered spaces. But game-type behaviors have been around for a long time. It's just something that we're now starting to take seriously for learning. On top of

that, that's the reality of life. Everything's changing. The way people communicate is changing. We've seen the way Facebook sort of has enculturated an entire world into building networks. Whether you think that's positive or negative is irrelevant. The reality is that that's what's going on. So I think that's something you have to take seriously, regardless of the complications involved. You have to be willing to publish, and say, hey, that's out of date. This is what we know now, and all of that body of knowledge, I think, leads into important research that we need to pay attention to and take seriously.

I understand that games are your focus right now, but do you think that there are any other important directions within CALL that you might want to pursue in the future?

Yeah, absolutely. I mean, anything that involves collaboration and social learning. Kind of Web 2.0, but I think we've moved a little bit beyond that now. But mostly I think CALL people need to start really paying attention to social change and the things that are coming about as a result of some of those changes. Goodfellow, Lamy, and Steve Thorne talk a lot about some of these critical approaches to using technology. I tend to agree with them and say yes, language learning is changing, but the needs of language learners are also changing, and that's really what we need to pay attention to as we're moving forward in CALL research in general. Telecollaboration, for example. There's a new volume that just came out that addresses some of these issues, and I think that that's important and something that we really need to critically think about as we're moving on.

Thinking about students in SLA or SLS, in general, what sort of advice would you give to students right now? And which areas do you think are exciting to work in now?

My first piece of advice is something I always tell my advisees: find something you like doing. If you don't enjoy it, don't do it just because I think it's cool as your advisor. Don't do it because it's the cool thing to do, or it's what's going to get you published, or it's what's going to get you a job. Of course those are important considerations, but in the end, it's about what you're interested in working on and what you find most compelling because it's too much work and too much sacrifice not to enjoy what you're working on. I'm lucky in that I found a career and a profession that I really enjoy. It's hard work, and it's tiring, and all of those things, but it's worth it in the end because you feel like you're at least moving forward. In terms of CALL, again, I think there are tons of things. My advice is, find a language learning problem that you want to solve because there are tons of them out there. There are lots of things we don't know. Of course, I have to put a plug in for pragmatics because I think it all boils down to pragmatics in a lot of ways. Steve Thorne and I were talking the other day, and he mentioned the idea that everything boils down to how communication happens. So for me, of course, that's an interesting area, but it's also really complicated. It doesn't have a lot of black and white answers, and so, if there's a student who really needs black and white, statistical responses, it's harder to find in pragmatics. You have to be willing to tolerate ambiguity to get the kind of answers we're looking for. But what are you interested in solving and what tools are available out there to

help us solve that, be it replicating different mechanisms or utilizing what's already out there?

Thank you very much for your time today.

Thanks for inviting me, and I'm glad to be here at MSU.

Differences in English Compliment Responses Between Native English Speakers and Chinese English Learners

Wen-Hsin Chen

PhD student, Second Language Studies Program

Michigan State University

chenwen4@msu.edu

This research discusses the compliment responses of Chinese-speaking English learners in the United States by using the conversation analysis methods. I compared the examples collected for this study to Pomerantz's (1978) data for compliment responses made by native speakers of American English. The participants were sixteen Chinese ESL learners. The primary result was that some compliment responses used by Chinese-speaking English learners were similar to those of native speakers of American English. Most of the time, Chinese-speaking English learners chose to accept the compliments with or without referent shifts. Even though they sometimes chose to disagree with the speaker, they would incorporate other-than-you references in their expressions. Nevertheless, Chinese-speaking English learners used "really?" or "oh really?" as one of their responses, which is quite different from American English.

Speech acts are often analyzed in terms of function—for instance, requesting, refusing, apologizing, and complimenting. Second language (L2) learners display noticeably different L2 pragmatic competence (defined as language users' knowledge of communicative action and how to carry it out appropriately according to context) than native speakers, both in production and comprehension (Bardovi-Harlig, 2001). One possible explanation is that learners may hesitate to transfer L1 strategies that they are not certain are universal. A second possibility is that they mistakenly transfer strategies that they wrongly assume to be universal (Kasper & Schmidt, 1996). Speech acts reflect cultural values. If second language learners fail to perceive cultural meanings behind surface syntactic meanings, the potential for conversation

breakdowns or miscommunication increases.

Pragmatic transfer, an area related to the development of pragmatic competence, was defined as the use of first language (L1) speech norms in the speaking or writing of an L2 (Wolfson, 1989). Pragmatic transfer consists of two main categories: positive and negative transfer. Positive pragmatic transfer occurs when an L2 learner successfully conveys his/her intended messages by transferring conventions of usage shared by L1 and L2 (Al-Issa 1998; Kasper, 1992); in contrast, negative pragmatic transfer refers to L2 learners' use of their L1 speech norms in inappropriate L2 contexts (Felix-Brasdefer, 2004), thus committing pragmatic failure—failing to understand interlocutors' intentions. Thomas (1984) pointed out that pragmatic failure is much more serious than linguistic errors.

That is, learners who make linguistic errors seem to be thought of as less proficient language users, whereas those who fail to conform to the target language pragmatic norms tend to be considered unfriendly or impolite (Bardovi-Harlig, Hartford, Mahan-Taylor, Morgan, & Reynolds, 1991).

Cross-cultural aspects of speech acts have gained increased research attention over the past three decades (Cohen, 1996; Wolfson, 1981, 1989). Wolfson is considered one of several pioneers in research on cross-cultural distinctions regarding compliments. She observes that comments considered to be compliments by Americans can be viewed as insulting by people from other cultures. For example, an intended compliment made by an American speaker that an addressee looks unusually attractive can be viewed by French receivers as implying that the reverse is true (Wolfson, 1981).

Knowing how to compliment is important, but it is equally important to know how to respond to a compliment (Nelson et al., 1996). For non-native English speakers (NNEs) living in the United States, knowing how to respond to compliments appropriately is considered important because of the higher frequency with which Americans give compliments. Pomerantz (1978) has completed several studies on compliment responses in English-speaking countries, but little research exists on responses used by NNEs in ESL contexts. Yu (2004) investigated how Chinese EFL and ESL learners responded to compliments given by native English speakers. The results showed evidence of pragmatic transfer: A high percentage of rejection was found. Rejection of compliments is regarded as having good manners in Chinese culture but improper or even rude in American

culture. More research on compliment responses used by NNEs and pragmatic transfer is required.

In light of the above background, I set out to investigate similarities and differences in compliment responses between Chinese learners of English and American native English speakers. A central goal was to identify factors that can assist L2 instructors. Throughout this paper I will refer to four types of compliments: on appearance, on possessions, on ability, and on personality traits (Nelson, Al-Batal & Echols, 1996; Wolfson, 1989).

Compliment Responses

Pomerantz (1978) is responsible for two conversational principles governing compliment responses. The first states that most recipients agree with and/or accept compliments, and the second states that most recipients avoid self-praise. She classified realization patterns of compliment responses into two categories: (a) acceptance, which can be divided into the categories of appreciation tokens (Excerpt 1) and agreement, where acceptance tokens are followed by agreement components (Excerpt 2), and (b) rejection, including disagreement (Excerpt 3). For acceptance, appreciation is preferred over agreement. Rejections are routinely associated with disagreement.

Excerpt 1: Acceptance (Pomerantz, 1978, p. 84, Excerpt 4)

- A: Well-I-I wannid to say I enjoyed your class so this morning, and too.
- B: Well, thank you.

Excerpt 2: Agreement (Pomerantz, 1978, p. 85, Excerpt 8)

- A: Oh it was just *beautiful*.
- B: Well *thank* you Uh I thought it was quite nice...

Excerpt 3: Disagreement (Pomerantz, 1978, p. 87, Excerpt A)

- H: Gee, Hon, You look nice in that dress
- W: Do you really think so? It's just a rag my sister gave me.

Excerpt 4: Evaluation shift (Pomerantz, 1978, p. 94, Excerpt 33)

- B: I've been offered a full scholarship at Berkeley and at UCLA.
- G: That's fantastic
- B: Isn't that good

Excerpt 5: Evaluation shift (Pomerantz, 1978, p.100, Excerpt 46)

- K: Those tacos were good!
- B: You liked them...
- K: I loved 'em, yes.
- B: I'm glad, but uh, next time we have 'em we'll, the uh, the tortillas a little bit more crispy...

Although the preferred response is for the recipient to accept a compliment with what Pomerantz calls an *appreciation token*, this conflicts with the self-praise avoidance constraint. There are two solutions to this conflict: *evaluation shifts* and *referent shifts*. Evaluation shifts are praise downgrades, with recipients using evaluative descriptors that are less positive (Pomerantz, 1978.). Recipients can do this by either agreeing (e.g., Excerpt 4) or disagreeing (e.g., Excerpt 5) with a compliment, using scaled-down or more moderate praise terms.

Referent shifts include compliment responses in which "recipients of praise proffer subsequent praises of other-than-self referents" (Pomerantz, 1978, p. 107). In Excerpt 6, F credits K by focusing on K as the weaver of the blanket in question. Later in the sequence, F admires the blanket with a positive assessment. K initially responds with a scaled-down agreement (line 10), and then with a credit shift away from herself as the weaver (line 12).

Excerpt 6: Referent shift (Pomerantz, 1978, p. 102, Excerpt 2c, p. 103, Excerpt 2d)

- 1 F: ..What ayou making?
- 2 K: It's a blanket
- 3 F: Did you weave that
- 4 [yourself
- 5 K: [I wove this myself
- 6
- 7 D: She wove [all of this herself
- 8 F: [Ya kidding
- 9 F: That is [beautiful
- ↓10 K: ['N that nice
- 11 R: Yah. It really is.
- ←12 K: It wove itself. Once it was set up

In the Manes and Wolfson (1981) corpus (cited in Wolfson, 1989), other-than-self referents occur at high frequencies. In Excerpt 7, A responds to S's compliment by shifting the credit to her mother.

Excerpt 7: Referent shift (Manes & Wolfson, 1981)

- S: That's a pretty sweater.
- ← A: My mother gave it to me.

Excerpt 8: Referent shift (Pomerantz, 1978, p. 107, Excerpt 52a)

- E: You lookin good
- G: Great. So'r you.

A second type of referent shift is returning compliments, as in Excerpt 8. According to Pomerantz (1978), returns are most frequent in interaction openings and closings.

In summary, Pomerantz (1978) observed that "praise downgrades are prevalent subsequent to compliments with other-than-you references incorporated, and appreciation tokens show a priority positioning over agreements and disagreements" (p. 108)

The Present Study

The motivation for the present study is to investigate similarities and differences in compliment responses between Chinese learners of English and American native English speakers. Toward this goal, I compared the examples collected for this study to Pomerantz's (1978) data for compliment responses made by native speakers of American English.

Method

Participants

The sixteen Mandarin-speaking participants (eight males and eight females) were all enrolled at a large Midwestern university in the United States in Spring 2008; two were exchange undergraduate students and the rest were graduate students. These ESL learners were all native speakers of Mandarin Chinese from Taiwan. All of them had received English instruction before coming to the United States. All had minimum scores of 213 on the CBT.

Procedure

Data were collected by audio recording individual interviews between the researcher "K" who is a female Taiwanese graduate student in her 20s and each participant in a language lab

for approximately 3 to 5 minutes. Each interview started with some demographic questions (e.g., age, major, hometown, and length of U.S. residence), followed by free conversation. During the course of conversation, the researcher initiated compliments according to the four above-listed categories (i.e., compliments on appearance, possessions, ability, and personality traits) and later analyzed participant responses. At the end of each interview, each participant was asked whether he/she realized the purpose of the interview. No participant reported any awareness of the focus of the study.

Data Analysis

Audio files were transcribed by the researcher. All data were transcribed using Jefferson's notation system (Atkinson & Heritage, 1984) and subjected to conversation analysis (see Appendix A).

Results and Discussion

Upon receiving possession compliments, the Chinese participants tended to accept them, but the same was not true for ability compliments, which generally elicited disagreement. Responses to appearance compliments were mixed between acceptance and disagreement. Only one instance of a personal attribute compliment was noted; the recipient disagreed at first and then incorporated an other-than-self referent and evaluation shift.

Similarities

The main similarities between compliment responses made by the Chinese learners of English in this study and native speakers of American English in Pomerantz's study were acceptance of prior compliments and the incorporation of referent shifts and/or evaluation shifts into responses. Some

Chinese learners accepted compliments by using appreciation tokens such as “thanks” or “thank you” in the same manner as Americans. In the conversation leading into Excerpt 9, K wants to know the required qualifications for becoming a TA like S. S responds by saying that applicants need to pass a speaking test. K compliments S on his English speaking ability, and S responds with an appreciation token.

Excerpt 9

- 1 K: Wow, that’s why your
2 speaking is so good.
•3 S: Oh, thanks.

Pomerantz (1978) notes that “appreciation tokens and agreement components may be used in combination and/or as alternatives by recipients” (p. 86)—a phenomenon also observed among Chinese learners of English. In Excerpt 10, K gives two compliments, and Hu responds differently to each: with appreciation to the first and agreement to the second, adding an appreciation token both times.

Excerpt 10

- 1 K: I like your sweater,
2 [it looks] great=
•3 Hu: [thank you]
•4 Hu: =yeah, thanks.

Referent shifts

Some participants incorporated referent shifts in their responses to compliments. In Excerpt 11, K learns that B has drawn a picture and asks B to show it to her. B finds the drawing and gives it to K, and K compliments B on her drawing ability. B responds with

both acceptance and agreement, and then shifts the credit to her mother.

Excerpt 11

- 1 K: Wow, so cute, [wow]
2 B: [thanks]
3 K: I like your drawing.
4 B: Thank you. Yeah
5 K: And, so have you learned
6 <how to draw before?>
7 B: No, I didn’t. I just, you know,
8 draw whatever I want and
9 then it turns out to be like
10 this.
11 K: Uh huh. Yeah I think your
12 skill is quite good.
13 B: Thanks. (.) Yeah. I think it’s
14 (.) probably: I have
•15 this:(.) >I don’t know< ability
16 from my mom?

Other Chinese participants used referent shifts when accepting compliments on appearance. In Excerpt 12, L shifts the credit for the shirt he is wearing.

Excerpt 12

- 1 K: I- I think you look great in this
2 shirt.
3 L: Look great?
4 K: Yeah, looks more younger=
5 L: =Oka[y
6 K: [no, much younger
•7 L: Only when I wear this shirt?
8 K: I don’t [know
9 L: [hh h [.h
10 K: [hh h .h but I
11 think, really=
12 L: =[okay, thank y[ou.
13 K: [your [style is good-
14 great.

Referent shifts were also observed in responses to compliments on possession.

In Excerpt 13, R accepts K's compliment and then shifts the credit to her mother.

Excerpt 13

- 1 K: Oh, this is your pajamas!
2 R: Yeah.
3 K: I like it. It looks so goo:d. hh
4 h .h
•5 R: Thank you. My mother got
6 this from um: Indonesia.

In addition to other-than-self referents, the Chinese learners of English in this study also returned compliments in the same manner as Americans in Pomerantz's study. In Excerpt 14, M accepts K's compliment and then returns it to the speaker.

Excerpt 14

- 1 K: Wow, nice jacket, you look
2 great in it.
•3 M: Thank you very much (1.0)
4 you too. hh h .h

Evaluation shift plus referent shift

As mentioned above, praise downgrades with other-than-you references are prevalent in American English (Pomerantz, 1978); I found that the Chinese participants in this study also applied this strategy. In Excerpt 15, Y is a new exchange student, and K compliments him on his English speaking ability. Y responds with disagreement using an other-than-you reference, shifting the credit to his need to get a passing score on the TOEFL. Next, he downgrades the credit to the level of "a little" English speaking ability.

Excerpt 15

- 1 K: This is- this is your first time

- 2 in United States?
3 Y: Yeah.
4 K: but I think your English is
5 very good.
•6 Y: Oh, no. uh: .hh uh: but we
7 need to pass the TOEFL
8 exam when we want to::
9 >exchang- get exchanged<
10 here
11 K: Uh [huh.
12 Y: [So, uh: but I think (.)
13 that's- that's why I have
•14 a:- a little English speaking
15 ability, I think. hh h .h

Another example of this kind of response is given in Excerpt 16, in which K and C discuss the GRE and K discovers that C got a very high score. K credits C for a personal trait (intelligence), and C shifts the credit to hard work. When K reasserts the compliment, C shifts the credit to her memory, and then downgrades that same characteristic in the last line.

Excerpt 16

- 1 K: You must be very smart.
•2 C: No, but I- I memorize so lot
3 of vocabulary be- before
4 I actually take the test.
5 K: But I still think you're very
6 smart=
7 C: =hh h .h w[hy,
8 K: [because you get
9 high on GRE s- test
•10 C: Um. I don't know, maybe
11 memory is great.
12 K: [hh h .h
•13 C: [hh h .h But now I- I don't
14 remember at all hh h .h

In the final excerpt for this section, H accepts K's compliments on her appearance, and then downgrades the credit by saying the jeans she is wearing

are cheap (specific problems related to this compliment category will be discussed in a later section).

Excerpt 17

- 1 K: I think your- your jeans looks
- 2 great, you look great in you-
- 3 this jeans.
- 4 H: Oh, thank you. hh h .h (1.5)
- 5 It's really (.) cheap.

Differences

The primary differences in compliment responses between the Chinese learners of English who participated in this study and native speakers of English in Pomerantz's study were the use of agreement, rejection, and "Really?" / "Oh, really?" as compliment responses.

Agreement

Pomerantz (1978) argued that agreement in American English is less frequently expressed than appreciation, and that agreement has more restrictions on production—that is, most agreement comments emerge from second assessments “which are systematically altered relative to the prior compliments, containing scaled-down or more moderate praise terms than the priors” (p. 94). Results from the present study conflict with Pomerantz's. In addition to finding several instances of agreement among the Chinese participants, I noted that they seldom used a second assessment with scaled-down terms, preferring instead to use such words as “yeah” to express agreement (see Excerpts 10 and 11 above). In Excerpt 18, K tastes a bowl of soup cooked by H, and compliments H on her cooking ability. H responds with agreement in the form of “yeah” and reconfirms the compliment with “I like to cook.”

Excerpt 18

- 1 K: Wow, it's so good
- 2 H: Really? hh h .h
- 3 K: Yeah, you must very- be very
- 4 good at cooking
- 5 H: Yeah. I like to cook.

Rejection

According to Wolfson (1989), simple disagreement with a compliment is less frequent among American native English speakers, who prefer to show disagreement by downgrading. I noted several strong examples of praise downgrades among my Chinese participants (Excerpts 15, 16, and 17) and fewer instances of simple disagreement. In Excerpt 19, K hears from her friends that P's application has been accepted by one of Taiwan's top four universities. P rejects K's compliment of her ability with an explicit “No.”

Excerpt 19

- 1 K: You have been accepted by
- 2 Chiao Da? =
- 3 P: =Yep.
- 4 K: Wow, you are so great
- 5 P: No: hh h .h

Excerpt 20 is another instance of a compliment for ability. W wins a scholarship funded by the Taiwan National Science Council and receives a compliment from K. W rejects K's compliment with simple disagreement rather than downgrading.

Excerpt 20

- 1 K: So you must be outstanding.
- 2 W: Uh:: .hh (0.8) hh uh:: I don't
- 3 know, but, in their point of
- 4 view, maybe, but I don't

5 think so.

Some of the participants showed the same tendency for compliments on appearance. In Excerpt 21, K compliments F on being slim, which is considered a positive value among Chinese. F rejects the compliment by saying “no” three times with rising intonation. This is very different from Pomerantz’s finding (1978) for native English-speaking Americans, who tend to respond by reducing the compliment. Furthermore, Pomerantz noted that Americans tend to challenge or disagree with rejections and reassert their compliments. Note that in Excerpt 21, K insists on the compliment two times, but F keeps rejecting it.

Excerpt 21

- 1 K: I just want to say even
2 though you come- you come
3 to [A- America for half- um:
4 F: [um hum
5 K: for eight months you always
6 look so slim, yeah.
•7 F: ↑ No, no, no, that’s not true.
8 Because you know, I have
9 gained weight for five kilos
10 so far
11 ((unintelligible sounds))
12 K: But compared to me, you-
13 you really look slim I think,
14 and I envy you.
15 F: hh[h .h
16 K: [hh h .h
•17 F: No, you didn’t see the fat
18 here you know, here
19 ((F showed her fat to K))
20 F: hh[h .h
21 K: [hh h .h
22 F: You know
23 K: But I really think so.
•24 F: No: I don’t think so.

Chen (2003) asserts that Taiwanese are more likely than American native English speakers to use the Chinese language equivalent of “Really?” or “Oh, really?” to respond to compliments given by other Taiwanese. I found that the Chinese participants in this study transferred this strategy to their second language—an instance of what Wolfson (1989) calls pragmatic transfer: the use of L1 speech norms when speaking or writing an L2. Schegloff (2007) describes the interjection “oh” as a “change-of-state token” used to mark or claim information receipt, and “Really?” (with or without a preceding “oh”) as a request for further information. Schegloff also writes:

Several of the turn types which can be used for other-initiation of repair can also be used to mark some utterance or part utterance as of special interest, and worthy of further on-topic talk. Among these forms are repeats or partial repeats, “pro-repeats” (such as “He is?”), and “really,” all with or without a preceding “oh” ... “Oh really?” marks the answer as “news,” and provides for further expansion of its telling. (2007, p. 155, 157)

Chinese participants in this study used “Really?” and “Oh, really?” in their responses to compliments on appearance, possession, and ability. In Excerpt 18 above, after K tastes some soup cooked by H and compliments it, H responds by saying “Really?” After K provides more information, H agrees with the compliment. In Excerpt 22, A responds to K’s compliment with “Oh, really?”, K provides more information, and A finally accepts the compliment.

Excerpt 22

- 1 K: I like your hairstyle. I think

- 2 this one fits you.
 •3 A: Oh, re- really?
 4 K: Yeah, you look so cute, yeah,
 5 very cute. hh h .h
 6 A: Okay, thanks.

In contrast to Shegloff's finding, I noted that my Chinese participants sometimes used "Really?" or "Oh, really?" as markers of received information, thus requiring no further response. In Excerpt 23 (a compliment of possession), E responds to K's "I like your jacket" comment by saying "Oh, really?" E does not wait for K's response, but goes on to express his acceptance of K's subsequent compliment, "It looks so great" with a thank you.

Excerpt 23

- 1 K Wow, I like your jacket,
 2 [it looks so great
 •3 E [>oh, really?<
 4 E Thank you.

Excerpt 24

- 1 K: Oh, I just think your English
 2 is quite goo:d. It's very fluent
 3 and native- like [I think]
 •4 Z: [hh .h]
 Really? Thank you. hh h .h

Another example of this finding is Excerpt 24, an instance of a compliment on ability. Again, Z does not wait for K's response to his "Really?", but instead continues to express his acceptance of K's compliment.

Pedagogical Implications

The study findings can be used to assist language teachers and learners. For teachers, they can support efforts to develop teaching materials for compliment/compliment response

sequences; since they provide knowledge of what authentic Chinese conversations sound like and identify two kinds of response strategies that should be taught: (a) appreciation tokens over agreement and disagreement, and (b) praise downgrades and/or other-than-you references when expressing agreement or disagreement with compliments.

For Chinese learners of English, the study results will support their perceptions of similarities and differences in compliment responses between them and native speakers of American English. The data can also help them understand what kinds of response strategies they need to learn in order to respond to compliments in culturally appropriate ways. The findings can reinforce the idea that when expressing agreement or disagreement with compliments, it is proper to incorporate praise downgrades and/or other-than-you references in their responses. In addition, Chinese learners of English can also benefit from learning that the phrase "Oh really" performs a different function for native English speakers than it does for Chinese.

Suggestions for Future Study

Pomerantz (1978) never explicitly categorized compliment responses according to compliment type, and described few instances of responses to compliments for reasons other than personal attributes; in contrast, I only collected one example of this type of compliment. My plans are to collect a much larger body of data on responses to all four compliment types mentioned in the Introduction section, for both native speakers and Chinese learners of English.

I found several instances of two compliments being made within a single turn, thus making it difficult to categorize compliment responses. For example, in Excerpts 14 and 17 it was difficult to tell which compliments the recipients were responding to. In both cases I categorized them as responses to the second compliment, but future researchers may find other ways to distinguish among multiple compliments and responses. Another topic that researchers may be interested in analyzing is the function of laughter, which was commonly heard in many of the recordings I made. Laughter may have some function in compliment/compliment response sequences that requires further analysis.

Finally, Pomerantz (1978) discussed how the compliment-givers she observed frequently reasserted praise when their recipients showed disagreement. However, she did not offer data about how addressees responded to reasserted compliments. My data suggests that addressees may change their responses when speakers insist on reasserting compliments (as in Excerpt 22) or they may be equally insistent in their responses (as in Excerpt 21). This competition is a topic for future study.

Conclusion

In this study I analyzed similarities and differences in compliment responses between Chinese learners of English and native speakers of American English. By and large, the Chinese participants followed Pomerantz's (1978) two conversational principles: "One is supportive actions, that is, responses which legitimate, ratify, affirm, and so on, prior compliments, and the other is self-praise avoidance" (p. 106). Most of the time the Chinese

participants chose to accept compliments with or without referent shifts. Despite occasionally disagreeing with a compliment-giver (especially when the compliments referred to ability), they incorporated other-than-you references in their responses. Praise downgrades were also observed in their compliment responses.

Unlike native speakers in Pomerantz's study, the Chinese learners of English in this study used "Really?" with or without a preceding "oh" as a common compliment response. Based on analyses of the use of the Chinese equivalent of "Really?" in studies of Chinese speakers of Mandarin (e.g., Chen, 2003), this strategy is considered a L1 transfer. Americans use "Really?" in their conversations, but rarely as a compliment response. Americans may perceive a Chinese learners' use of "Really?" as an invitation for further information, whereas Chinese consider it a culturally acceptable compliment response—a potential scenario for cross-cultural miscommunication.

References

- Al-Issa, A. (1998). *Sociopragmatic transfer in the performance of refusals by Jordanian EFL learners: Evidence and motivating factors*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9823511)
- Atkinson, J. M., & Heritage, J. (1984). *Structures of social interaction: Studies in conversation analysis*. Cambridge, England: Cambridge University Press.
- Bardovi-Harlig, K. (2001). Evaluating the empirical evidence: Grounds for instruction in pragmatics? In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 13–32).

- Cambridge, England: Cambridge University Press.
- Bardovi-Harlig, K., Hartford, B., Mahan-Taylor, K., Morgan, B., & Reynolds, T. (1991). Developing pragmatic awareness: Closing the conversation. *ELT Journal*, 45, 4–15.
- Chen, S. (2003). Compliment response strategies in Mandarin Taiwanese: Politeness phenomenon revisited. *Concentric: Studies in English Literature and Linguistics*, 29(2), 157–184.
- Cohen, A. D. (1996). Speech acts. In S. McKay & N. Hornberger (Eds.), *Sociolinguistics and language teaching* (pp. 383–420). Cambridge, England: Cambridge University Press.
- Félix-Brasdefer, J. (2004). Interlanguage refusals: Linguistic politeness and length of residence in the target community. *Language Learning*, 54(4), 587–653.
- Kasper, G. (1992). Pragmatic transfer. *Second Language Research* 8(3), 203–231.
- Kasper, G., & Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition*, 18, 149–169.
- Manes, J., & Wolfson, N. (1981). The compliment formula. In F. Coulmas (Ed.), *Conversational routine: Explorations in standardized communication situations and prepatterned speech* (pp. 115–132). The Hague, the Netherlands: Mouton.
- Nelson, G., Al-Batal, M., & Echols, E. (1996). Arabic and English compliment responses: Potential for pragmatic failure. *Applied Linguistics*, 17(4), 411–432.
- Pomerantz, A. (1978). Compliment responses: Notes on the cooperation of multiple constraints. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 79–112). New York, NY: Academic Press.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis I*. New York, NY: Cambridge University Press.
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 31–112.
- Wolfson, N. (1981). Compliments in cross-cultural perspective. *TESOL Quarterly*, 15, 117–124.
- Wolfson, N. (1983). An empirically based analysis of compliments in American English. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 82–95). Rowley, MA: Newbury House.
- Wolfson, N. (1989). *Perspective: Sociolinguistics and TESOL*. New York, NY: Newbury House/Harper & Row.
- Yu, M. (2004). Interlinguistic variation and similarity in second language speech act behavior. *Modern Language Journal*, 88(1), 102–119.

Appendix A

Transcription Conventions

[the point where overlapping utterances begin overlapping
]	the point where overlapping utterances stop overlapping
=	linked speech
(o.o)	pauses or gaps in what is approximately tenths of a second
(.)	micropause
-	truncated word, self-editing marker
...	medium pause
:	an extension of the sound or syllables it follows
::	a longer extension of the sound or syllables it follows
italics	syllables stressed by amplitude, pitch and duration
•	draws attention to location of phenomenon of direct interest to discussion
> <	an utterance is delivered at a pace quicker than the surrounding talk
< >	Inaudible utterances
↑	marked rising shifts in intonation
.	falling intonation
,	maintained (continuous) intonation
?	rising intonation
!	an animated tone
hh	audible aspirations
.hh	audible inhalations
(())	used to specify “some phenomenon that the transcriber does not want to wrestle with” or some non-vocal action, etc.

Investigating Incidental Vocabulary Learning in Conversation Classes: A Qualitative and Quantitative Analysis

Ayman Mohamed
PhD student, Second Language Studies Program
Michigan State University
mohame44@msu.edu

This study examined incidental receptive and productive vocabulary gains within conversation class interactions. Sixteen Mexican learners of English attended four videotaped conversation lessons where 40 target words were incorporated into different types of exposure. Stimulated recall interviews with students highlighted the effect of cognates, learners' access to passive vocabulary, and use of their vocabulary knowledge in learning related words. Posttests revealed a correlation between frequency of exposure and receptive/productive gains. Mean scores showed that students most often learned task-essential words, followed by words mentioned with synonyms, and last, those mentioned without an explanation. A two-way ANCOVA revealed main effects for cognates, and a statistical interaction between cognate status and types of exposure to target words, and a moderate effect of frequency of mention on receptive knowledge. Results provide implications for ESL teachers who consider incidental learning of vocabulary within their conversation lessons.

In vocabulary acquisition reviews, there are usually references to distinctions between intentional and incidental modes of learning new words. Incidental vocabulary acquisition was defined by Wesche and Paribakht (1999) as what happens when learners are focusing on understanding meaning rather than on the explicit goal of learning lexical items. Gass (1999) maintained that incidental learning suggests reduced cognitive processing in that the learner does not exert that much energy to commit an item to memory as it is the case with intentional strategies of learning vocabulary. Hulstijn (2001, 2003) drew a methodological distinction that incidental learning occurs when learners are not told beforehand of an upcoming test after a given treatment. Based on

this distinction, one assumes that a typical conversation class, which mainly involves meaning-based communication—with no intention to teach vocabulary—can be considered an optimal setting for incidental learning.

Reviews of vocabulary studies usually indicate that incidental vocabulary learning is much rarer than teachers might like to think, and is often slower than explicit learning (Horst, 2005; Hulstijn, 2001; Macaro, 2003). Horst (2010), however, maintained that there are certain opportunities for incidental vocabulary acquisition in a communicative class and from the teacher's speech. Similarly, Nation (2001) posited that a vocabulary-learning goal can be integrated into speaking tasks to encourage incidental learning. The present study evaluates

the opportunities that conversation classes can afford for incidental vocabulary development by examining the factors of input and interaction that could encourage incidental intake and retention of new words while students are engaged in meaning-based interaction and speaking tasks.

The main route by which students are able to learn vocabulary incidentally in the classroom is through their interaction with the teacher and other students in the target language. The interaction hypothesis (Long, 1996) claimed that input becomes comprehensible through interaction. If there is a breakdown in communication, this actually helps learners notice gaps or deficiencies in their ability to communicate. They can subsequently try to repair these, and thus the process can facilitate language acquisition. Many studies have validated this theory and found evidence that interaction can lead learners to notice problems in their interlanguage and attain higher skills in the second language (Gass & Varonis, 1994; Mackey & Philp, 1998; Mackey, 1999; Pica, 1994; Pica, Young, & Doughty, 1987; Polio & Gass, 1998; Swain & Lapkin, 1998).

Following the assumptions of the interaction hypothesis, a number of studies have investigated the question of whether vocabulary acquisition occurs incidentally during interaction and meaning negotiation of target words. However, the majority of these studies have looked at the question only through controlled lab experiments. For example, Ellis, Tanaka, and Yamazaki (1994) found that in all cases, the group that was exposed to interaction showed the highest comprehension and acquisition scores of target words. A similar study by Ellis and He (1999) showed that negotiated input yielded better learning

than baseline input, but the kind of negotiation that allowed students to produce and modify their own output yielded the highest scores in comprehension and retention of target vocabulary. LaFuente (2002) set up an information-gap activity with participants divided into groups that received modified input, negotiated input, and pushed output. Results confirmed the general advantage of negotiation over the modified input in the scores for comprehension and receptive acquisition. Negotiation with pushed output had the advantage of higher scores for productive acquisition and retention.

Taken together, the results of these previous studies seem to indicate that, through a process of hearing and producing meaningful input and output in tasks where learners need to exchange information, vocabulary can be incidentally comprehended and acquired. However, if we are to know whether such acquisition can take place in the classroom, we have to look at situations that are not limited to negotiation, but that also involve spontaneous interaction where vocabulary is likely to occur naturally in different instances and contexts. One study that closely touched upon spontaneous interaction was Brown, Sagers, and LaPorte (1999). The authors investigated natural oral and written journal exchanges between a teacher and nine advanced EFL learners over a whole semester. A comparison was made between the nature of input in the oral and the written modes and the nature of output produced by students. The new vocabulary items produced and used by students after the teacher had used them were considered possibly acquired because of this interaction. The number of words acquired from oral

input was greater than that from the written input. There was better acquisition when students recognized their lexical gaps, meaning that they indicated that they did not know the exact word they wanted to use. Several exchanges on a single topic led to better acquisition, and the topics chosen by students yielded better results than those stemming from topics chosen by the teacher.

One study that has analyzed real classroom interaction and the acquisition of vocabulary was (Dobinson, 2001). In this study, teachers carefully prepared four lessons that included the target vocabulary that they would teach in their English class, and the sessions were videotaped. The author collected lesson plans from the teachers to record the vocabulary items they targeted and compare them to what students actually gained. The study found that learners recalled and retained words better when they were mentioned, focused on, or repeated within class interaction. The study intended to examine intentional vocabulary acquisition, but evidence of incidental learning was also found when learners recalled words that came up spontaneously in class and were not intended by the teacher.

A considerable bulk of research has been done on incidental learning of vocabulary through reading. The findings of such studies related acquisition to context clues, type of task, time on task, frequency, proficiency, or vocabulary size (Brown, Waring & Donkaewbua, 2008; Kweon & Kim, 2008; Paripakht & Wesche, 1999; Rott, 1999; Sanchez & Schmitt, 2010; Watanabe, 1997; Webb, 2008). However, there seems to be a lack of research on incidental vocabulary acquisition from oral input, particularly within classroom interaction. Horst (2010) conducted a

corpus-based appraisal of the opportunities that in-class teacher talk can afford for incidental acquisition of newly encountered words. Using vocabulary frequency profiles, the study considered factors of comprehensibility, repetition, and type of talk. The results suggested that attending to the teacher's speech is not an assured method of acquisition because important academic words and frequent words are unlikely to be encountered within the teacher's discourse exchanges with students in class. However, the study provided implications for the possibility of integrating these important words in meaning-based speaking tasks and activities.

Although few studies have looked at incidental vocabulary acquisition in the language learning classroom, the studies that have been conducted on incidental acquisition point to several factors that researchers should take into account when investigating this question. Ellis (1994) posited four factors that influence incidental vocabulary acquisition from oral input. He referred to these factors as intrinsic word properties, learner factors, input factors, and interaction factors. Ellis suggested that learners can also acquire vocabulary from noninteractional input through the various techniques of teacher-discourse, which include definition, conjunction, elaboration, apposition, and parallel structures. However, Ellis expressed concerns that most of these factors could just be ways that guarantee comprehension, but not necessarily acquisition. Gass (1999) summarized some of the more important intrinsic word factors by noting that a word is more likely to be learned incidentally if there are cognates between the L1 and L2, if a considerable number of exposures occur, or if a number of other

related L2 words are known. If none of these conditions hold, the learner then resorts to intentional learning techniques.

In a post-hoc analysis of research results, Schmitt (2008) pointed out that the tasks that were more effective for vocabulary learning in interaction studies (e.g., Ellis & He, 1999) and reading studies (e.g., Watanabe, 1997) were more engaging than other less effective tasks. In this sense, he highlighted the fact that engagement with vocabulary is the key for incidental learning, and that any intervention that makes target words essential in a task or a class activity would evoke more engagement with lexical items on the part of the learner.

Many researchers have used the classroom interaction setting in focus on form research, first introduced by Long (1991). Loewen (2005) and Nassaji (2010) investigated incidental focus on form in class interaction and both found evidence of the effectiveness of form-focused episodes (FFE) on the acquisition and development of target linguistic features, vocabulary being one of them. In their meta-analysis, Mackey and Goo (2007) found that interaction and feedback were more beneficial when the target features were lexical items rather than grammar items. In the present study, the concept of planned focus on form is used in the sense that vocabulary items are planned to be embedded into various types of exposure, and subsequent measures of retention are likely to point to the most effective modes of exposure in classroom setting.

Research Hypotheses

One relevant hypothesis of the study is that learners are likely to notice new words as they are mentioned in context without explanation, manage to guess

their meanings, and show retention of these words in a vocabulary posttest. This review shows several lines of support for hypothesizing that incidental learning can take place naturally in the classroom. Mackey, Gass, and McDonough (2000) found that learners were more likely to attend to lexical feedback more than to syntactic or phonetic feedback. In line with these results, Gass and Alvarez Torres (2005) investigated the different effects of attending to input and interaction on the acquisition of grammar and vocabulary. One relevant implication these researchers cited was that vocabulary required less attention and less externally driven focus because the learner's internal mechanisms are more helpful in attending to vocabulary as a non-complex and non-abstract area of language.

Another hypothesis is that students acquire words used in conjunction with or appositive to a synonym more easily than words encountered without explanation. Support for this hypothesis comes from Watanabe's (1997) study of written input, which found that words provided with synonym glosses or embedded in appositives in reading passages were learned better than words that were simply read as part of the context of the passages. This finding can be tested orally by looking at what happens when words are mentioned in context only, when mentioned with a definition or explanation, and when in conjunction with or appositive to synonyms.

A further hypothesis is that students can retain task-essential words in speaking activities more often than other non-essential words. Task-essentialness has been validated through earlier studies on the interaction hypothesis. Several studies on

vocabulary acquisition have found that when learners produce their own meaningful input and output through interaction tasks and engage in negotiation of meaning, they retain words better than just hearing native speaker input (Ellis & He, 1999; Ellis, Tanaka, & Yamazaki, 1994; LaFuente, 2002; Newton, 1995). Given that speaking tasks are common practice in second language classrooms, and particularly in conversation classes, it would be a further support for assumptions of task-essentialness to investigate tasks conducted naturally in the classroom rather than lab-controlled environments.

Research Questions

Based on the above hypotheses, the present study is intended to answer the following questions:

1. Does classroom interaction afford opportunities for learners to attend to novel words mentioned in context and show recognition and retention of their meanings?
2. Are words mentioned with synonyms or appositives more likely to be noticed and retained than words mentioned in context without explanation?
3. Do task-essential words yield better acquisition and retention than other non-essential words in the context of classroom interaction?
4. In general terms, what interaction factors afford more opportunities for incidental learning of new words as they occur in the classroom context?

Method

Participants

The participants for the present study were recruited from a pool of Mexican English as a Second Language (ESL) students who were at an intensive summer program at an American

university for a period of four weeks. Sixteen students (12 females and 4 males) consented to participate in the study. Their ages ranged from 19 to 28. They were placed in the intermediate and upper-intermediate proficiency level in English in their program.

Materials

Vocabulary checklist. Vocabulary items were drawn from the most and least frequent words in the academic word list, supplemented with other general use words. They were added to a checklist in which the students had to check the words that they knew. The test consisted of two hundred words. After participants completed the test, a total of 35 words that all the learners had checked as unknown were chosen to be embedded into the teaching sessions. Five additional words that spontaneously came up during the treatment were added to the analysis. The list of target words is given in Table 1 below.

Topic checklist. A checklist was prepared with suggestions of general topics that could be the focus of the lessons. Four topics that were preferred by all the participants were selected for the lessons. These topics were distributed in the sessions in the following order:

Session 1: Culture and concepts from our life

Session 2: Relationships and gender roles in different cultural views

Session 3: Pollution, weather change, and natural disasters

Session 4: Dreams, luck, and superstitions

Classroom context. Standard ESL textbooks, websites, and activities were used for the lesson plans. Four two-hour meetings were scheduled with the students to conduct the lessons after they signed the consent forms. The

Table 1

Target Words for the Sessions

bond	comprise	notion	attain	sustain
ethic	assess	perspective	assign	advocate
bias	priority	core	exceed	adequate
encounter	norm	deviation	devote	expand
confine	diminish	emerge	utilize	distort
isolate	assemble	vivid	ambiguous	intervene
reveal	widespread	exploit	precise	anticipate
chores	errand	conscript	strive	distort

sessions were videotaped using three video cameras, each camera capturing one group of three or four students, along with the teacher's work and talk. The target words were embedded in different ways during the sessions to test their effects. In one type of exposure, certain words were made task-essential, as the students would have to use them to complete tasks. Only a few words were intended to be explicitly defined or elaborated on, given the typical nature of a conversation class. In another planned type of exposure, certain words were intended to be mentioned in conjunction with or in apposition to a synonym. Examples of contexts in which a word is said in conjunction with a synonym can be drawn from the transcript of the sessions:

Today I want to expand or extend this a little bit by talking about our relation to our environment

An example of a word said as an appositive to a synonym or explanation is the following:

Yes, which means they could utilize, make good use of their own resources for...

The rest of the target words, which was most of them, were intended to be only mentioned in context without further explanation. This is due to the typical nature of a conversation class, where vocabulary does not usually show up as the focus of instruction. In addition to the target words, other words occurred naturally without being planned.

Because it was almost impossible to plan and control everything that happened in class, the video sessions had to be transcribed to investigate the types of exposure and frequencies of target words and relate them to recall and retention results. A complete chart was produced that contained the target words and described their occurrences, frequencies of mention, and how they were presented in class. This chart is shown in the appendix below.

Procedures

Conversation sessions. Students attended four class meetings on four successive days. Each session lasted about 2 hours and included warm-up activities, video or audio sections, topic discussion, and group activities. Students were arranged in groups before

each session started to allow for setting up the video cameras.

Testing.

Stimulated recall protocol (SRP). SRP is a methodological procedure that has been advocated by Gass and Mackey (2000, 2005), two leading researchers in the field of second language research. The researchers described SRP as an introspective method for collecting data. In this method, the participant is given a reminder of a specific situation so that mental processes used during this situation are stimulated and recalled.

For testing immediate recall of vocabulary, SRP sessions were held individually after each teaching session in a linguistics lab with five randomly selected participants. In each interview, the participant generally watched the video recording of a given session right after class or in the morning before the following class. The participant would watch the video segment of his or her group on a 17-inch laptop with headphones on. The participant was instructed to stop the video at any time to give a comment of any type about vocabulary, structure, or pronunciation points. No attempt was made to interrupt the silence of students while watching. The researcher asked questions only when the learner gave a comment. The questions were “Did you guess the meaning?”, “Did you write the word down?”, “Did you look it up in a dictionary?”, and other questions relevant to students’ comments.

Productive vocabulary test. This test consisted of fill-in-the-gap sentences. To avoid any ambiguity in picking the intended target word, the first two or three letters of each target word were provided in each sentence. The sentences were created by selecting the most frequent academic usage of

target words in the Corpus of Contemporary American English (COCA). Two points were given for a correct answer, one point for a semantically appropriate response (a word similar in meaning and appropriate for context), and zero for an incorrect or blank response.

Receptive vocabulary test. This test was intended to collect more self-reports from the participants about their own learning of the target words. It was adapted from the vocabulary knowledge scale (Wesche & Paribakht, 1996), which uses a 5-point scale from 1 (have never seen the word) to 5 (full familiarity and usage of the word in a sentence). The receptive posttest only contained the 40 target words, and was conducted right after the productive section. Both the productive and receptive tests were conducted in a meeting one day after the last teaching session.

Analysis. Qualitative data from the stimulated recall interview was used as an indicator of what should be expected in a quantitative post test. Major themes were coded from students’ responses, and a list of recalled words was used to identify what factors encouraged more recall of target words. The interviews brought up other factors that were interesting to add to the quantitative analysis besides the type of exposure. These were cognate status and frequency of mention of target words. Based on that, scores of receptive and productive knowledge were analyzed into two sets of two-way ANCOVA with cognate status (2 levels), type of exposure (3 levels) as the independent variables, and frequency of mention as a covariate in order to isolate its effect from the type of exposure variable. Post-hoc analyses were made when significant values were found.

Results

Qualitative Data

The stimulated recall sessions provided an initial view on immediate vocabulary intake after each class. On average, learners recalled 11 words from the target vocabulary over the course of the four sessions. For the type of exposure, it was found that the words recalled by the most participants were those words that had been made task-essential within the sessions. Examples are words like bond, chores, errand, vivid, and ambiguous. From the naturally-occurring non-target words, students also recalled words that were task-essential in different activities. Examples were words like surgeon, newscaster, flatter, and bargain. Students recalled words from the listening and video tasks that required filling information into a table or taking notes for discussion. Examples of these words were chipped, unfold, storage, conservation, and tremendous. For other text-based words that were less essential to meaning, students reported that they either consulted a dictionary or asked their classmates for a quick translation or explanation. Examples of these words were cashier, threat, measure, and efficient. On the other hand, some students reported that they encountered unknown words but they did not have to use the dictionary because they understood the whole meaning from context.

The least recalled words were the words that were just mentioned in context or in conjunction with a synonym. From these, only four words were recalled by any participant as newly-learned words in the stimulated recall sessions. These words were assess, assign, conscript, and intervene. However, students recalled words mentioned in context more easily if they

were cognates, a fact that was explicitly pointed out by some of the students. Examples were words like essence, decade, and inevitable. These words, however, were not targeted for the treatment, but they occurred naturally in teacher's talk, speaking tasks, or class discussion.

Students reported learning new words as a result of realizing their morphological relations to already known words. Some examples are like the following: perceiving the word distant to be related to the word distance, the word costly to be related to the known word cost, and the word pollutant to be related to pollution.

Interaction and exposure to vocabulary aided students to remember words, to access new meanings, to confirm their knowledge of partially known words, and to recognize meanings for words they had heard before but did not know understand. Examples of students' comments were statements like: "I remembered I studied this word long ago," "I remembered it in class," "I think I heard this word before but I did not learn it," or "I only know one meaning of this word, but I learnt it can have another meaning in a different topic." Students reported that they knew some words, but that it was the first time that they had heard them within a certain expression. Examples of these were the expressions culture shock or extended family and the phrase it has to do with, which a student reported to have guessed from context.

Quantitative Results

Scores in the receptive knowledge test were entered into a two-way ANCOVA with cognate status and type of exposure as independent variables, and frequency of mention as a covariate. Results showed a statistical main effect

for type of exposure, $F(2, 38) = 8.07$, $p = .02$, partial eta-squared = 0.37, and a statistical effect for cognates $F(1, 39) = 21.9$, $p < .001$, partial eta-squared = 0.45. The interaction between cognates and type of exposure approached significance, $F(2, 38) = 3.17$, $p = 0.058$, partial eta-squared = 0.35. The effect of frequency was significant $F(1, 39) = 32.17$, $p < .001$, partial eta-squared = 0.54. A Scheffé post-hoc test showed a significant difference between task-essential words and no explanation words ($p = .002$).

For type of exposure, descriptive statistics showed that task essential words were learned most often, followed by words mentioned with a synonym, and last, those mentioned with no explanation. Mean scores and standard deviations are shown in Table 2. A visual representation of receptive test is shown in Figure 1.

The same type of two-way ANCOVA was also performed on the dependent variable of productive gain. The main effect of type of exposure was statistically significant, $F(2, 38) = 12.57$, $p < .001$, partial eta-squared = .43, with high power (0.93). There was a significant effect of cognate status, $F(1, 39) = 18.7$, $p < .001$, partial eta-squared = .36. The effect of frequency was not statistically significant, $F(1, 39) = 1.03$, $p = .32$, partial eta-squared = .030. A Scheffé post-hoc test showed a significant difference only between task-essential words and no explanation words ($p < .001$).

Mean scores of productive gain showed that minimal learning took place during the treatment. Descriptive statistics of productive scores are shown below in Table 3 and are represented visually in Figure 2.

Overall, task-essential words yielded better gains both receptively or

productively. Words mentioned with synonyms were more salient to learners than words mentioned with no explanation. The factor that some words were mentioned more frequently than others- ranging between two and twelve encounters- had a moderate effect on how words were retained, but it was through interaction with other factors that results could be interpreted. The factor of cognates largely determined the percentage of learned words. When words were cognates, chances were higher that learners would guess and retain them under any type of exposure within class interaction. This was noticeable mainly in receptive knowledge while there was a very low productive gain overall, except when words were cognates and used essentially in speaking tasks.

Discussion

The main question of this study considered the assumption that a conversation class would afford opportunities for incidental vocabulary acquisition in a variety of different contexts, subject to a number of factors. Results provided initial implications for the significance of the type of exposure and word properties, especially cognate status. The factor of frequency of exposure as well was not a target variable, but the data that the study provided indicated that it was important to test it. A significant effect was shown for frequency of mention in receptive, but not productive gains. This comes in line with literature on the effect of exposure frequency on the quality of vocabulary acquisition (e.g., Folse, 2006; Horst, Cobb, & Meara, 1998; Webb, 2007). The more a learner encounters a novel word, the more likely it will be acquired and retained.

Table 2
Descriptive Statistics for Receptive Acquisition

	<i>N</i>	<i>Mean</i>	<i>SD</i>
No explanation			
Cognates	18	2.33	2.91
Noncognates	7	0.71	1.11
Total	25	1.88	2.62
With synonym			
Cognates	4	4.50	4.65
Noncognates	5	2.80	2.77
Total	9	3.55	3.57
Task-essential			
Cognates	3	10.66	.58
Noncognates	3	3.00	3.60
Total	6	6.83	4.79

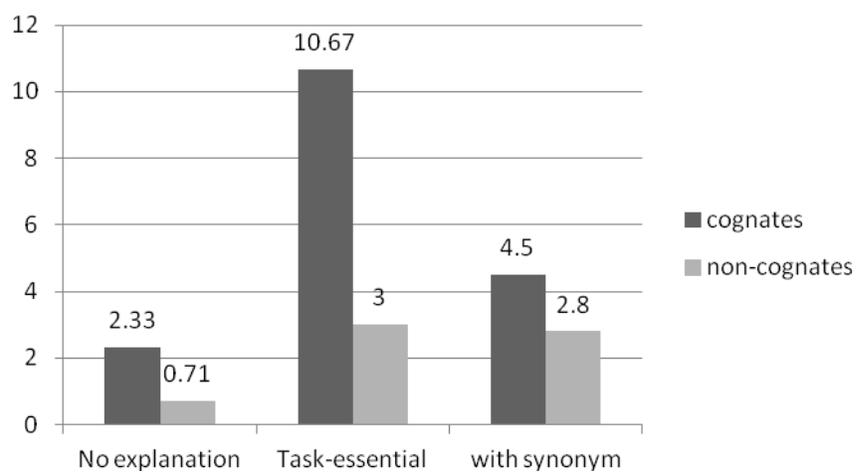


Figure 1. Receptive vocabulary learning.

Table 3
Descriptive Statistics for Productive Acquisition

	<i>N</i>	<i>Means</i>	<i>SD</i>
No explanation			
Cognates	18	0.50	0.86
Noncognates	7	0.43	0.53
Total	25	0.48	0.77
With synonym			
Cognates	4	.25	.50
Noncognates	5	1.20	1.64
Total	9	.78	1.30
Task-essential			
Cognates	3	7.0	2.65
Noncognates	3	0.67	1.15
Total	6	3.83	3.92

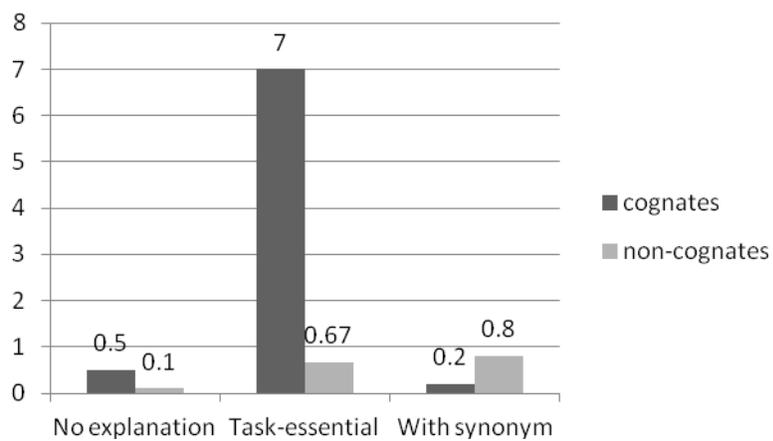


Figure 2. Productive vocabulary learning.

An important concern, however, was the possibility of having productive gains in vocabulary from a conversation class and the factors that could determine the feasibility of this finding. Overall, receptive gains were considerably higher than productive gains. The results from the productive test support the conclusions of the earlier studies on interaction and vocabulary acquisition, since most of the produced words were task-essential cognates. This implies that only the words that learners actually used in class and were more intrinsically salient were more likely to move from the level of recognition to the level of production.

One specific hypothesis of the study was that students in a conversation class can notice new words as they are mentioned in context without explanation, infer their meanings, then retain them in a posttest. This hypothesis was not confirmed. The least frequently acquired words were the ones mentioned in context without explanation. This could imply that students did not notice these words because they did not hinder comprehension, and thus learners did not need to pay attention to these new words, or that there was not enough context to guess the meanings of words. In support of this finding, Laufer (2005) raised a strong case in favor of form-focused learning, claiming that learners, by comprehending the overall message, are less likely to pay attention to individual words. If learners do so, the chances are not high that they will guess meanings correctly, especially if they do not know 98% of the discourse, and thus no considerable incidental learning would be expected in this case. In light of the noticing hypothesis (Schmidt, 1990), initial learning or intake takes place only when the learner notices the

word and the relationship between its form and meaning, and thus attention is involved, even if it is only through an incidental process.

An insignificant exception within the results for the 'mentioned in context' type of exposure came from individual learners when one or two participants learned some of these words. These students reported having written down the words and checked them in a dictionary. This suggested individual vocabulary learning motivation on the part of these learners. In this sense, it seems that this does not fit into the concept of incidental learning. By noticing the word, writing it down, and checking it in a dictionary, the learners were intentionally trying to add a new vocabulary item into their lexicons. The amount of deliberate attention and intentional focus does not seem to be classified as incidental learning. This calls for further research on learner strategies and self-reports of incidental learning and the fine line between intentional and incidental learning conditions.

Another hypothesis of the study was that students could notice words mentioned in conjunction with or apposition to synonyms, recognize their meanings, and show retention of these words in a posttest. This hypothesis was initially supported. The mentioned with synonyms type of exposure yielded a significantly increased rate of acquisition. These words were shown to be more salient for the students and yielded better vocabulary gains than those for which the words were mentioned in context. This makes sense because some students were able to find the relationship between words and their synonyms. More studies are required to investigate whether oral input could be more or less supportive

of incidental learning of vocabulary in different contexts.

A further hypothesis was that students would retain task-essential words better than other words. Task-essential words yielded significant vocabulary recall and retention rates within the treatment. This can be explained in terms of salience. When students had to use the words for the completion of tasks, words became salient to them and were more likely to be retained. The set for learning was further facilitated when these task-essential words were cognates. Task-based interaction was validated here as a factor related to vocabulary gain, as has been found in experimental studies. The more students had to use the words, the more they were likely to retain and produce them in a later test.

Concerning word properties, cognate status was a significant factor in reception and production. This implies that cognates in this study actually facilitated learners' access to new words and improved performance. For reception, learners showed better performance in all types of exposures when words were cognates. In production, learners did almost the same on words mentioned with synonyms and words mentioned without explanation, but a difference emerged when the word was task-essential. Further studies need to address the factors that encourage noticing and recognition of cognates, which could involve frequency of mention, closeness of cognates, proficiency, aptitude, phonological awareness, or meta-cognitive language skills. Ellis (1999) referred to cognates as the learner's potential vocabulary, but research needs to explore how learners perceive cognates in different contexts, from oral input as well as written input.

Qualitative results from the stimulated recall sessions provided further support for factors of incidental learning and introduced additional factors that need to be explored in later research. Students recalled words that occurred naturally within classroom interaction without being targeted for the treatment. Learners recalled words that were close cognates in addition to being task-essential, as well as the words that teachers explained explicitly. The listening tasks that required focus or filling of gaps yielded recall results for newly-learned words. Students also recognized words that were morphologically related to already known words. They were aided by interaction in class to remember meanings of words, to access new meanings, and to confirm their knowledge of partially known words. This could be explained in terms of access to the learner's passive vocabulary, which would include those words that the learners had already encountered, but were not internalized as part of their active or productive lexicon. Most of these results are in line with what Gass (1999) proposed about the factors that encourage incidental learning: frequency of exposure, cognates, and knowledge of related words. All these factors facilitate the process of making certain words salient to learners so that the minimum amount of attention required for incidental learning takes place (Ellis, 1994; Gass, 1999; Hulstijn, 2003).

Conclusions

The present study has provided preliminary observations about the nature of incidental vocabulary learning within a real conversation class situation. It has introduced a primary idea that incidental learning in conversation

classes is possible, to a certain extent, under certain conditions. ESL teachers can consider this factor when they prepare their lesson plans for conversation sessions under the assumption that these lessons can be used for practicing speaking and communication as well as providing opportunities for new learning. The teacher should also consider the possibility that surfaced in the study: that students can be aided by interaction to access their passive vocabulary, remember meanings of words, or discover new meanings. This, in turn, may gradually enhance the spoken proficiency of learners by moving passive vocabulary items from perception to the realm of production through the teacher's incidental revisiting of partially known words on the part of the learners.

However, the implications provided by the study do not undermine or ignore the importance and efficiency of intentional learning of vocabulary because it was not hypothesized that the participants in this study would necessarily retain the acquired words over longer periods. As Nation (2001) posited, productive learning of vocabulary has to do with repeated exposure and practice. Schmitt (2008) maintained that incidental and intentional learning approaches are complementary and that they require one another. A question not yet answered is what the ordering effects of incidental exposure and explicit focus would be on the quality of vocabulary acquisition.

Limitations and Future Directions

A major limitation of the study is the small sample size, which makes the study close to a pilot experiment that provides observations and directions for a wider-scale and longer-term research. The study could also count as a case

study in that it studied Mexican students in particular in an ESL context with Spanish as the L1. This variable particularly brought up the effect of cognates and examined its significance. Other contexts with different L1s and in other ESL or EFL settings are likely to reveal other aspects of a conversation class and its interaction patterns.

The goal of the study was to replicate a naturalistic classroom setting while controlling, as much as possible the way several words were presented and used in class. The application of this methodology was less controlled than anticipated. The distribution of the types of exposure did not show equal numbers of words in each category. Cognate status was not controlled either because target words were randomly selected based on a pretest checklist. Additionally, the target words for the study focused on a selection from the academic word list, which does not cover the typical corpus of naturally occurring vocabulary in class activities (Horst, 2005) or their frequency distributions. Further research is required to refine the methods and designs that should be used in natural classroom research.

The present study did not measure levels of proficiency because students were placed at the intermediate level in their program. The question arises whether there is a threshold of proficiency that qualifies learners to notice and acquire new words in conversation classes and thus dispose them to boosting their lexicons from natural sources. Another promising area of research would look at aptitude and individual differences in phonological short-term memory and how these relate to the quality of incidental vocabulary acquisition. Along similar lines, vocabulary size measures were

hypothesized to be a predictor of subsequent learning of new lexical items.

Certain methodological challenges are usually involved in natural classroom research.

Some of these challenges surfaced as limitations for this study. Working with a larger sample for longer class hours could reveal more factors and effects in the long run. Recording more classroom data and coding different instances of vocabulary-focused conversation also seems to be a promising area to investigate incidental learning. Theoretical and empirical accounts are still needed towards a more refined operationalization of the distinction between incidental and intentional modes of vocabulary acquisition. Further research would be interesting to follow incidental learning from natural exposure to novel words, as opposed to classroom interaction.

References

- Brown, C., Sagers, S. L., & LaPorte, C. (1999). Incidental vocabulary acquisition from oral and written dialogue journals. *Studies in Second Language Acquisition*, 21(2), 259–283.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136–163.
- Dobinson, T. (2001). Do learners learn from classroom interaction and does the teacher have a role to play? *Language Teaching Research*, 5(3), 189–211.
- Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 35–61). Amsterdam, The Netherlands: John Benjamins.
- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam, The Netherlands: John Benjamins.
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2), 285–301.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, 44(3), 449–491.
- Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273–293.
- Fuente, M. J. d. l. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24(1), 81–112.
- Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), 319–333.
- Gass, S. M., & Alvarez-Torres, M. J. (2005). Attention when? An investigation of the ordering effect of input and interaction. *Studies in Second Language Acquisition*, 27(1), 1–31.
- Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.
- Gass, S., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, 19, 3–17.
- Gass, S., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition: Introduction to the special issue. *The*

- Modern Language Journal*, 82(3), 299–307.
- Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL*, 41, 87–106.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61(3), 355–382.
- Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language*, 22, 161–180.
- Horst, M., Cobb T., & Meara, P. (1998). Beyond *A Clockwork Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, England: Cambridge University Press.
- Kowen, S., & Kim, H. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language*, 20(2), 191–215.
- Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition*, 27, 361–386.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). San Diego, CA: Academic Press.
- Macaro, E. (2003). *Teaching and learning a second language: A review of recent research*. London, England: Continuum.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557–587.
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22(4), 471–497.
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 433–464). Oxford, England: Oxford University Press.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82(3), 338–356.
- Nassaji, H. (2010). The occurrence and effectiveness of spontaneous focus on form in adult ESL classrooms. *The Canadian Modern Language Review*, 66(6), 907–933.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, 11(2), 159–177.
- Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2

- vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21(2), 195–224.
- Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, 44(3), 493–527.
- Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, 21(4), 737–758.
- Polio, C., & Gass, S. (1998). The role of interaction in native speaker comprehension of nonnative speaker speech. *The Modern Language Journal*, 82(3), 308–319.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21(4), 589–619.
- Sánchez, P. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, 22(1), 31–55.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363.
- Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, 82(3), 320–337.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20, 232–245.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19(3), 287–307.

Appendix

Target word occurrences and results

Amount and Type of Exposure vs. Acquisition

Word	Frequency	Type of exposure	Recalled in SRP by	Retained receptively by	Produced by
ambiguous	3	Task-essential	8	11	5
priority	3	Task-essential	4	11	4
vivid	4	Task-essential	9	10	5
perspective	9	Mentioned in context	1	10	1
norm	3	In conjunction with a synonym	0	9	0
notion	3	In apposition with a synonym	0	8	0
chores	12	Task-essential + class discussion	2	7	5
bias	9	Defined and elaborated	0	6	4
utilize	4	Mentioned in context	2	6	0
adequate	3	Mentioned in context	1	6	0
assess	3	In conjunction with a synonym	4	5	4
exceed	3	Mentioned in context	3	5	0
intervene	4	Mentioned in context	1	4	3
assign	4	Mentioned in context	1	4	2
sustain	4	Text-based/not task-essential + mention	0	3	1
emerge	1	In conjunction with a synonym	0	3	0
reveal	1	Mentioned in context	0	3	0
errand	9	Task-essential + class discussion	0	2	5
core	2	in conjunction with a synonym	0	1	0
conscript	4	Mentioned in context	0	1	1
encounter	3	Mentioned in context	0	1	0
deviation	1	Mentioned in context	0	1	0
devoted	3	Mentioned in context	0	1	0
confine	3	Mentioned in context	0	1	0
widespread	2	Mentioned in context	0	1	0
bond	5	Task-essential	0	0	5
precise	1	Mentioned in context	0	0	0
comprise	3	Mentioned in context	0	0	0
ethics	1	Text-based/not task-essential	0	0	0
attain	3	Text-based/not task-essential + mentioned	0	0	0
advocate	2	In apposition with a synonym	0	0	0
disrupt	2	Mentioned in context	0	0	0

Word	Frequency	Type of exposure	Recalled in SRP by	Retained receptively by	Produced by
strive	3	Mentioned in context	0	0	0
expand	1	In conjunction with a synonym	0	0	0
distort	1	Mentioned in context	0	0	0
exploit	1	Text-based/not task-essential	0	0	0
diminish	2	In conjunction with a synonym	0	0	0
isolate	2	Mentioned in context	0	0	0
assemble	1	Mentioned in context	0	0	0
anticipate	1	Mentioned in context	0	0	0

Review of *Talk Time Student Book 2: Everyday English Conversation*

Stempleski, S. (2007). *Talk Time 2: Everyday English Conversation*. Oxford, England: Oxford University Press.

Review by Yunson Shin
Alumna, MA TESOL
Michigan State University
padao102@yahoo.co.kr

Talk Time is an American ESL book for promoting communicative proficiency for beginning to intermediate students. Interactive and communicative activities such as information gap tasks promote the fluency in collaborative activities and task-based language learning. Interactive activities induce pushed output, the negotiation of meaning, and allow students to develop automaticity. Extensive communicative activities help students use grammar rules implicitly. This book is a good source for promoting communicative skills, yet supplementary reading materials could help to activate schema and improve language learning. Additional activities such as pretask activities and cultural content could help make language learning more effective.

Talk Time is one of three levels in an American English ESL series. The book contains topics on everyday life for high-beginners to low-intermediate ESL students focusing on speaking, vocabulary, and grammar. Each unit includes five main sections: speaking, listening, grammar, conversation, and communication tasks. There are also three sub-sections: "Memo," "Extra," and "Helpful Language." In combination, the sections promote communicative proficiency effectively, as well as the internalization of grammar rules; however, replacing some of the conversational examples with additional resources and topics would help to promote speaking proficiency even more effectively.

The textbook proceeds as follows: The main vocabulary for each unit is introduced. Then, the speaking and listening sections develop

communicative proficiency through examples of conversation. For example, Unit 1 contains names of jobs, including those of teacher, dentist, and tour guide. The "Extra" subsection enables more vocabulary practice with topics in model dialogues through activities. The sessions focusing on communication e-tasks contain interactive activities like surveys, games, and information gaps. For example, the information gap activity in Unit 11 is an interactive pair activity about shopping items presented using two different pictures.

The speaking and listening combined skills promote, along with topical vocabulary, speaking proficiency for beginning-level language learners, helping them to build fluency and confidence in preparation for real-life situations. The listening training facilitates internalized learning of vocabulary for everyday life topics like

jobs, feelings and emotions, weekends or vacations, shopping, hobbies, travel plans, and cooking.

The section on grammar provides students with ways to use rules in communicative contexts and functional expressions in communicative tasks. For example, students plan weekend activities with partners, meeting the definition of a successful task: that it be related to the use of real-world language and resemble natural discourse (Ellis, 2000).

Speaking activities can improve language proficiency by encouraging students to negotiate meaning in collaborative learning settings, where they communicate with partners in pairs or groups. Along these lines, Watanabe and Swain (2007) explained peer-peer collaborative dialogue as an important mediator of L2 learning. As a result, conversations encouraged through vocabulary, listening, communicative tasks, and interactive activities guide students to express their everyday life experiences while building confidence.

Communicative tasks and activities in this book draw upon theoretical concepts from task-based language teaching (TBLT) and the output hypothesis. TBLT promotes pushed output through information-gap tasks and other communicative activities, bringing a real-world language focus to the negotiating of meaning. Izumi (2003) suggested that pushed output allow students the opportunity to develop automaticity, and Gass (1988, as cited in Izumi 2003) added that output plays a role in the dynamic, correlated acquisition process, when comprehended input is moved into intake through language production (Izumi, 2003). These activities encourage students to talk freely about past experiences and current needs in

everyday life by expressing likes, dislikes, making requests, and giving advice. Conversations in real life situations and communicative tasks provide students with integrated focus-on-form practice. Conversations introducing basic grammar points help students internalize grammar rules implicitly, following Ellis's (2006) suggestion that the construction of basic knowledge is needed for a form-focused approach so that students can use and extend their existing knowledge in a meaning-focused approach. Ellis (2006) also implied that a curriculum with integrated communicative tasks provides extensive focus-on-form grammar lessons. Skehan (2003) suggested that the use of extensive communicative activities is needed for learners who struggle to use language, as opposed to explicit learning about language. Students can then internalize grammar rules in communication tasks introducing basic grammar structures and fulfilled in conversational contexts.

The strength of this book is in how it builds communicative proficiency through topical vocabulary lessons and the introduction of grammar points through communicative contexts. In addition, contextualized real-life situations such as looking for directions on a street map facilitate internalizing grammar rules through unconscious learning with real-life photos of cities, people, and maps. Laufer & Hulstijn (2001) described how tasks and materials that interest students contribute to increasing these students' motivation.

Additional readings and listening tasks including authentic readings could help compensate for the book's weaknesses in written conversations, which do not provide sufficient authentic input opportunities. However,

a positive point is that authentic reading materials in the form of material-directed, guided opportunities will help students become more involved in the cognitive process and give students the benefit of facilitating targeted L2 acquisition (Williams, 2005). Another drawback to this volume is that the listening content, vocabulary, and details of conversation models are not demanding enough for intermediate learners to express themselves fully. Vandergrift (2004) stated that written support of listening material is not available in authentic and real-life listening situations. For this reason, I suggest that the addition of supplementary reading material sections would help students to activate schema, introducing background knowledge into conversation, which could be helpful for language learning. Adding a separate grammar section is not desirable because grammar points presented in the traditional non-contextualized manner are difficult to recall in a communicative context (Yang & Lyster, 2010). Conversational models and explicitly presented grammar forms could prevent students from noticing the gap in the use of language. In addition, each section provides rare opportunities to gain cultural understanding. Unit 2 contains only one cultural point, indicating that asking someone's age in some countries is impolite. More opportunities to understand cultural points could enlarge cultural knowledge in language learning. Also, adding pre-tasks would help to activate speaking proficiency. Williams (2005) insisted that increased pre-task planning time may have influence on fluency in students' L2 production when the amount of time and the proficiency of the learners are considered. Pre-tasks could allow learners to recall previous

knowledge to activate schemata and help language learning.

In sum, this is a generally effective ESL textbook that promotes communicative ability and the internalization of grammar rules. Students can practice and learn English through real-life, topic-based vocabulary learning, communicative tasks, and listening practice. With a few minor additions, such as adding the suggested topics and task resources, this book would be more effective.

References

- Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, 4, 193–220.
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83–107.
- Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics*, 24 (2), 168–196.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Skehan, P. (2003). Focus on form, tasks, and technology. *Computer Assisted Language Learning*, 16(5), 391–411.
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3–25.
- Williams, J. (2005). Form-focused instruction. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 671–691). Mahwah, NJ: Erlbaum.
- Watanabe, Y., & Swain, M. (2007). Effects of proficiency differences and

patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners. *Language Teaching Research*, 11(2), 121–142.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past-tense forms. *Studies in Second Language Acquisition*, 32(2), 235–263.

Multiple Perspectives on the Role of English Literacy Skills in the Curriculum of Young English-as-a- Foreign-Language (EFL) Learners in a Taiwanese Urban Context

Research proposal by Magdaleen Corne Lotter
Graduate, MA in TESOL program
Michigan State University
cornelot@gmail.com

This qualitative study explores the perceived influence of literacy skills on speaking skills for young English learners in Taiwan. Reading and writing skills are often neglected for the sake of improving learners' oral production. Previous research has shown a significant connection between oral production and literacy skills. The written modality is an asset in the EFL environment where there is a lack of native oral input, and integrating written language with oral for young ESL learners could lead to gains in oral proficiency. Interviews with teachers, school managers, parents, and a curriculum writer may indicate the inequality of time allocation towards literacy skills, and classroom observations may confirm this. Findings may reveal that teachers are not aware of parents' needs for their children's literacy skills. Teachers hope to spend more time on literacy in the classroom but are hindered by a full curriculum. Directions for future research are discussed.

Literacy skills are traditionally considered secondary to the development of oral language in many contexts (Harklau, 2002; Williams, 2008). Young learners of English commonly acquire the spoken language first, then they learn to read, and lastly, they write. From this view, one might infer that children's reading development could be influenced by his or her oral ability, and how well a child writes may depend on both his or her speaking and reading abilities. This has resulted in a discussion among language acquisition researchers and among language teachers over whether written English (reading and writing) should be used to develop oral English skills for beginning ESL students. In other words, is an adequate level of oral language skill

a necessity for literacy instruction in English? Researchers in child and adult language development (e.g., Kim, 2008; Olson, 2002; Williams, 2008; Wong, 2001) have indicated an interrelated and complex relationship between literacy and oral skills. Several researchers believe that there is no unidirectional influence of one modality over the other but rather a bidirectional relationship between written and oral modalities.

The aim of this qualitative study is to explore the perceived influence of reading and writing on the spoken skills of young English learners in Taiwan. In my experience as a language teacher in Taiwan, reading and writing skills are often neglected for the sake of improving young learners' oral production in Taiwanese private schools

despite the fact that previous research has shown a significant connection between oral production and literacy skills. In fact, integrating written language with oral production for young ESL learners might lead to greater gains in oral proficiency (Blake, 2009; El-Koumy, 1998; Kim, 2008; Weber & Longhi-Chrilin, 2001).

To understand what takes place in this context, I will examine the perspectives of some stakeholders at a school in Taipei, namely, teachers, school managers, parents of Taiwanese learners (of L2 English), and curriculum writers. The topics to be addressed in this study include the participants' views on teaching literacy skills, methods of teaching literacy, time allocation to reading and writing, time allocation to oral practice, and views on the links between literacy skills and oral production.

EFL students at the kindergarten and elementary levels are underresearched populations throughout the world, and research with regard to the development of second-language literacy among children scarcely involves such populations. Although this study does not focus on the development of speech and writing of learners within the EFL context, it does nevertheless investigate the focus of instruction, whether on oral or written English, and the possible reasons for preferring to focus on one modality rather than the other. It also investigates whether stakeholders are aware of the potential bidirectional relationships between the oral modality and written modality.

Literature Review

Second language teaching and learning has historically been about the acquisition of spoken language; in other words, the focus has been on teaching speaking because written production

seemed less likely than spoken language to be a reflection of English proficiency. Research, especially bilingual research, has concerned itself primarily with the study of spoken language (Leki, 2000; Valdés, 1992). Only a few empirical studies have investigated the effects of modality (Polio, 2012; Weissberg, 2006; Wong, 2001) and very little research has been done on the early L2 writing of young learners (Matsuda & De Pew, 2002).

The development of written and oral skills are often viewed as separate processes (Strube, 2011). Most researchers on child reading development hold the assumption that the development of reading depends on prior phonological awareness, and as such, literacy acquisition depends on a child's speech processing skills (Tarone & Bigelow, 2005). However, other researchers in child language development lean toward the opposite position in that the development of literacy increases phonological awareness. Olson (2002) stated that writing introduces our speech to us; that is, writing shows our speech as having a particular structure. "To segment words, the child has first to learn that an utterance can be segmented into words, and that knowledge too may be acquired in the process of becoming literate" (Olson, 2002, p. 156).

Some research challenges the idea that ESL learners need to become proficient in spoken English to learn the basics of written English. Several studies have taken for granted that children have basic implicit knowledge of their first language and thus a foundation for acquiring the form and use of another language in print as well as in speech (Weber & Longhi-Chrilin, 2001). In Taiwan, the situation is different in that knowledge of the learners' first language

(Mandarin Chinese) does not necessarily provide children with a foundation for L2 literacy because of the different writing systems used by Mandarin and English. The use of different orthographic systems has also received little attention in L2 literacy research, particularly with children. Buckwalter and Lo (2002) studied a five-year-old Taiwanese learner of English, and the case study gave insights into the debate as to whether the introduction of literacy in languages with two different writing systems helps or hinders literacy development in both languages. They found that their participant acknowledged Chinese and English as separate writing systems with different characteristics—and he was aware of the differences. These researchers also concluded that literacy development in one language had a positive effect on literacy development in the other. Interacting with text and constructing meaning from it led to foundational concepts in literacy. Reading and writing both English and Chinese help to develop the basic concepts of literacy. “This knowledge serves as a support base for literacy in any language, regardless of the surface level differences that may occur due to the nature of the writing system” (Buckwalter & Lo, 2002, p. 287).

The effect of home literacy practices on children’s language abilities and later academic success has been well documented (for understanding associations between early reading and later language skills, see Karass & Braungart-Riker, 2005). Other research has provided evidence that joint writing activities (writing activities completed through parent/child cooperation) were more effective for literacy development than joint reading for children aged three to five (Levy, Gong, Hessels, Evans,

& Jared, 2006). These joint writing activities improved children’s performance on phonological awareness and word writing. On the other hand, the usefulness of L2 reading for receptive skills was suggested by Elley and Mangubhai (1983). They found that reading skills transferred not only to productive skills (i.e., speaking and writing), but also to other areas of academics.

Kim (2008) argued that oral language and literacy skills can develop simultaneously. She provided two different types of instruction (i.e., integrated and oral-language-based instruction) to two young ESL learners. The results showed that the participant who was exposed to the integrated instruction made gains on most English oral and written assessment measures. Not only do these findings suggest that it is possible to develop literacy skills without a predetermined level of oral skills, but also that literacy skills can be used to develop oral language skills for young ESL learners. The findings showed that reading and writing can play a positive role in the development of oral language and that students’ reading and writing were important and might provide learners with chances to record their ideas as well as to further their language development.

Another relevant study was conducted by Weber and Longhi-Chirilin (2001). These researchers studied two Spanish first graders and suggested that children can readily apply themselves to reading and writing in English in spite of limited spoken ability. Both children achieved much toward acquiring early English literacy, such as reading words orally, without a strong oral foundation. These children, however, found themselves in a setting that allowed them access to spoken English most of

the day, which is much different in an EFL environment such as Taiwan, which will be discussed later.

A number of other studies have also shown that writing can improve oral ability (Blake, 2009; El-Koumy, 1998; Kim, 2008). El-Koumy (1998) used dialogue journal writing as a tool in the EFL classroom in Egypt to help improve oral fluency. The posttest results indicated that the experimental group that used dialogue journals scored significantly higher than the control group on oral fluency tests.

Blake (2009) addressed the issue of improving oral fluency in a second language with the use of internet chats. His study was conducted in an effort to contribute to research with regards to the oral-written connection. The significantly higher gain scores in oral assessment of the internet chat group in a university-level ESL class support the notion that oral fluency improvement is possible within a text-based environment. In an exit survey, parents, teachers and learners indicated their skepticism about the use of writing and reading to improve oral fluency, and therefore more studies are needed to promote the idea of reading and writing as important factors in oral fluency. Related to Blake's findings, Hardison (2011) found that the percentage of time L2 English (L1 Korean) graduate students spent using English (vs. their L1) for various types of electronic communication significantly predicted their fluency scores in an oral interaction task.

Several researchers have shown that the written modality could be helpful to draw learners' attention to form, and that could have a facilitative effect on overall proficiency. Van Patten (1990) indicated that adult L2 learners of Spanish have difficulty simultaneously

attending to the meaning and form of aural input, especially when the grammatical form is not essential for understanding the content. Van Patten only addressed the aural mode in his Spanish L2 data, but Wong (2001) compared the written and aural modes and focused on French learners' acquisition of English. Wong (2001) found that learners can pay attention to form and meaning at the same time in writing, unlike speaking. Even though the participants in the Wong study vary greatly from the participants in the current study (i.e., college level students vs. young learners), Wong's findings have relevance in that they clearly indicated that "attentional constraints do not affect the aural and written modes in the same way" (Wong, 2001, p. 360). Processing written input may be less taxing on the language learner's attentional resources because written input is segmented and can be reread.

The mutual interdependence of writing and oral skills is perhaps obvious, but in the past, speaking was seen as the precursor, and writing was viewed as the outcome of proficiency. Rubin and Kang (2008) suggested several ways in which written language acts as a foundation for oral proficiency. Acquiring the print code affects the metalinguistic representation of speech; that is, when "children can visualize language because they have cracked the print code, they consequently become more aware of the stream of speech as composed of segmentable units" (p. 215). While speaking can often stimulate writing, the opposite is also true. Learners may talk about their writing processes, or they may talk about their texts as objects. Writing can also script oral performance, or it can guide interaction. Writing requires a slower rate of production, and therefore, it

allows the opportunity for more reflection and revision (Rubin & Kang, 2008).

The acquisition of the ability to decode an alphabetic script has been shown to change the way in which an individual processes oral language (Tarone & Bigelow, 2005). In a study with illiterate adults, the results indicated that the acquisition of the grapheme-phoneme correspondence in learning to read an alphabetic script provided important cognitive tools, for instance, the awareness of linguistic units encoded in written language, for the processing of oral language. Tarone and Bigelow (2005) stated that an adequate SLA model should also be able to account for the learning experiences of illiterate and low-literate multilinguals, and the directionality between phonological awareness and literacy development cannot be fully understood by working exclusively with children. Thus, incorporating research with an illiterate adult population has the potential to give a much broader picture of SLA.

Other research has indicated that language may emerge first in the written modality before speaking (Harklau, 2002; Weissberg, 2006). The written modality took preference over the spoken modality as the preferred mode for the development of L2 syntax for a group of ESL learners at an American university (Weisberg, 2006). Certain grammatical forms appeared in particular modalities for all five participants in a variety of oral and written language production tasks, such as oral interviews and written essays. Irregular verb forms, personal pronouns, prepositions, and plurals most often appeared first in speech. Regular past morphemes, negatives, modal auxiliaries, passives and perfect verb

tenses appeared in writing before they appeared in speech. These findings also have pedagogical implications for the EFL and ESL classroom, and research with more participants could shed light on these findings. With these ESL learners, written English syntax appears not to have developed on the basis of an existing oral proficiency. These findings have several implications for L2 writing and speaking instruction in that they suggest a preference for writing over speech as the main modality for morphosyntactic development. Weisberg (2006) showed the importance of writing in the L2 acquisition process of adults, but some generalizations, to a limited extent, can be made with young learners in Taiwan because the situation is similar to what Weisberg described. Weisberg put it very aptly: "It seems clear that the L2 composition classroom is not just a place to learn about writing; for some students it may be the best place to learn the new language" (2006, p. 52).

The relationship between the development of written and oral proficiency is a dynamic and complex one. Williams (2008) discussed the influence of writing on the development of oral proficiency. Research has shown that writers are more likely to develop their writing when they have a chance to talk about it. Learners can also use the written modality to test out new forms and access acquired forms they do not yet totally control. The use of a new form in writing "increases the likelihood that it will be produced later in a more spontaneous setting, such as conversation" (Williams, 2008, p. 13). There is less pressure in writing than speaking, which allows learners a safe and more private place to try out new language about which they do not feel confident. The aforementioned

discussion makes apparent the possible benefits from the written modality for learners of English.

The role of writing and reading in lower-level curricula needs to be reconsidered. For example, Maxim (2002) concluded from his study of beginning L2 German learners at a university that they benefitted from a curriculum where extensive reading was incorporated. He also proposed that beginning students could develop more than just reading skills, but also greater grammatical and communicative competence. Even though Maxim's study involved adult beginning learners, it might be applicable to young learners in Taiwan.

The education system in Singapore has similarities with that of Taiwan in the sense that children are attending English schools, but they do not come from English-speaking homes. In addition, teachers in Singapore face many external constraints such as rigid syllabi and limited curriculum time (Ng & Sullivan, 2001). Moving away from a curriculum that relied heavily on writing, the Singaporean government implemented a Reading Skills Project (REAP) that focused on the acquisition of reading skills. Several years later, tests revealed that REAP schools outperformed non-REAP schools with regards to speaking skills, amongst others. Ng and Sullivan (2001) found that the students who read more also spoke English more confidently and responded more in classroom discussions.

English Within the Taiwanese Context

It is relevant to consider how the Taiwanese perceive the English language. English has been considered a prestigious language for study in Taiwan since the end of the Ching dynasty

(1644–1911), when the Chinese society started to be more welcoming of Western civilization. The prestigious position of English continued after World War II because of the ties between the United States and Taiwan (Wang, 2000). English, especially American English, has remained popular in Taiwan, and the Taiwanese government has promoted English education to a great extent in recent years. Wang (2000) indicated that English serves an instrumental function in Taiwanese society, in that Taiwanese people depend on English for knowledge from professional publications and English language media. Despite the popularity of English, Chinese remains the medium of instruction in both elementary and high schools, and both students and teachers indicated “that the language most often used in English class in high school is Chinese” (Wang, 2000, p. 129). Lai (2009) also confirmed that English, although a major foreign language taught in school, is not used much by people in society. In the city of Taipei, the only suburb where learners of English might be exposed to English in everyday life is Tienmu, a popular area for expatriates and their families. However, in most other areas of Taiwan, only Taiwanese, Mandarin Chinese, and Hakka are spoken, and learners of English do not have many opportunities to hear English outside of the classroom.

Children start learning English very young, as early as kindergarten. There is also a lasting trend for parents to send their children to private language schools or bushibans to better compete with peers and do well on entrance tests to be admitted to good elementary and high schools in Taiwan. Previously, English language education began at the secondary level, but since 2001, English instruction has been introduced at the

elementary level. This trend is occurring not only in Taiwan, but in other East Asian countries such as Japan and Korea. In the past, critics noted that language instruction focused too much on grammar and translation, with the result that students often acquired insufficient communication skills (Butler, 2004, 2007). To rectify this, the Taiwanese government began introducing English language education at the elementary level, with a particular emphasis on developing oral skills. The government provided several general guidelines for teaching English. To develop students' communicative abilities in English, the government suggested to teachers a number of activities such as games, songs, chants, and role plays (Butler, 2005). English is taught as an academic subject for around 72 lessons per year (40 minutes per lesson). The objectives articulated by the central government are "a) To develop students' basic English communicative abilities; b) To develop students' interests in and ways of learning English; c) To increase students' awareness of native and foreign cultures and customs" (Butler, 2004, p. 248).

The Taiwanese government also suggested that all English classes be conducted in English with a relaxed and interactive instructional method. Speaking and listening are the primary focus, and according to the government policy, "reading and writing should not be neglected" (Butler, 2004, p. 249). At the elementary school level, not many native-speaking teachers (NSTs) teach English, and English language instruction is usually done by individuals who have obtained English-related degrees or individuals who possess sufficient English proficiency based on the computer-based TOEFL

test. These teachers, however, often have insufficient proficiency to teach English effectively (Butler, 2004).

The Chinese culture of learning in Taiwan warrants some discussion. Taiwan is a highly exam-oriented society, and success on writing tests is usually a precondition for academic study (Chien, 2011). According to Yu (2008), emphasis is placed on memorization and analytical ability, rather than functional use of language for communication (see also Lai, 2009). The role of the teacher is the "source of knowledge," and Yu (2008) reported that Chinese teachers of English often have concerns about adopting Western approaches such as communicative language teaching. These sentiments were also supposed by Butler (2005) and Wang (2000). Wang found that grammar-based practices still reign in English classes in Taiwanese high schools. One of the reasons could be the way Taiwanese students learn Chinese. Chinese language learning is seen as the memorization of words and grammar. In Chinese, children learn to write first before reading, and it is presumed that Chinese learners should learn written words by writing them so that they can read them later (Hsu, 2004).

The same procedure is not applicable when Taiwanese students learn English because they learn to read first, and "most Taiwanese students will have only two years English composition writing experience at their 11th and 12th grades" (Hsu, 2004, p. 2). Longhi-Chirilín and Weber (2001) also noted that writing is not a regular practice in the ESL classroom.

According to Yang (1999), students in Taiwan have strong beliefs about becoming skilled in listening and speaking skills. Students believe that the purpose of studying English is to have

native-like speaking proficiency. Wang (2000) noted that pronunciation (specifically American English pronunciation) is an important factor in the Taiwanese context. Wang also reported that most English learners in Taiwan considered excellence in pronunciation to be the most important factor in improving English communication. Because the Taiwanese government emphasizes oral communication in their elementary school English curricula, Butler (2004) discovered that Taiwanese teachers felt that they needed a more balanced proficiency level across all skill domains, not only speaking. Butler (2005) also found that many Korean and Taiwanese teachers questioned the government's current policy and commented that students "need to have instruction in written English to facilitate their learning" (p. 437).

When EFL kindergarten students make the transition from private language schools to elementary school in Taiwan, they are faced with a variety of difficulties. The two education systems of kindergarten and elementary schools are quite different. In kindergarten, the students are used to an environment with small English classes, native-speaking English teachers who use different teaching methodologies, and curricula that focus on spoken skills. When they go to elementary school, they become part of classes with more than 40 students each. They also have Taiwanese teachers who do not follow the same teaching methodologies as the native-speaking teachers. At the elementary school level, there is also a focus on written English, rather than spoken English. Both learners and parents often complained to me that kindergarten did not prepare them sufficiently for the writing activities

done at elementary school. Similar frustrations of children entering first grade without much experience in literacy were found by Weber and Longhi-Chirilin (2001) and Harklau (2000). Students considered "good students" or "model students" often experience difficulties in elementary school, and they often rebel against the system and long for their kindergarten days.

I lived and worked as an English teacher in Taiwan for nine years, and during this time, I noticed that the learners went to elementary school with inadequate literacy skills, which had a detrimental effect on their experiences there. Parents often complained to me that their children did not want to study English anymore once they entered elementary school. I became interested in the topic because I wanted to make sure that these young learners were prepared for elementary school. In my experience, the written modality was neglected in the kindergarten classrooms at private language schools, and when these children went on to elementary school, they could not cope with the writing that was expected of them at that level. As a teacher, I was explicitly told by school managers that parents just expected their children to be able to speak English. In several conversations with parents, I received contradictory requests. Parents were very worried about their children going to elementary school without sufficient writing and reading abilities. To prepare these learners for elementary school, teachers might have to focus more on the written modality. In implementing this study, one should keep in mind the positive effect that writing can have on oral skills as well as the reasons for teachers and managers neglecting reading and writing at a young age;

therefore, I suggest interviewing teachers, parents, and managers. To triangulate the data, classroom observation data can be used.

Research Questions

This research is qualitative in nature. Issues to be addressed by this study include the participants' views on teaching literacy skills; their opinions regarding suitable materials; methods of teaching literacy; time allocation to reading and writing; time allocation to oral practice; and views on the links between literacy skills and oral production. This led to the following research questions:

1. What are the participants' views on teaching literacy skills?
2. What are the teachers' methods of teaching literacy?
3. What percentage of time is allocated to reading, writing, and oral practice in class?
4. What are the participants' views on the links between literacy skills and oral production?

Method

Research Site

The research site is a well known private school in Taipei, Taiwan. The school is very prestigious and attracts top students from all over the island. This language institute is the largest in Taiwan, with several branches in Taiwan and in countries such as Korea, Canada, and Singapore. The company employs native English speakers to teach English to Taiwanese learners aged three to sixteen. Classes are often co-taught by a NST and a Chinese teacher. The main office supplies all branches with the curricula (including books, audio CDs, props, and artwork) written and published by company employees.

These young learners have long school days that vary from 8 to 10 hours.

They are smart and highly motivated learners. By the age of six, they have studied Mandarin, English, Japanese, and French. The teaching philosophy of the company focuses on whole-child development and what the company calls "educare." Whole-child development implies that children will achieve the best results cognitively, emotionally, physically, and socially when they develop a balance of these intelligences. The curriculum has, therefore, been written with the methodology of simultaneously integrating several ways of learning. Educare is an approach to schooling that recognizes the indivisible relationship between educating and caring for a child; that is, if a child feels safe and cared for, then learning will come effortlessly.

These schools have three levels of classes: little (children aged 3–4), middle (children aged 5), and big (children aged 6–7). The classes that are relevant to this study are big classes, and all teachers interviewed should be big class teachers.

Participants

The participants for the study will be 10 native English-speaking teachers, five school managers, three parents of Taiwanese learners, and one curriculum writer. The employees at the time of data collection will either be working at one of the schools in Taiwan or at the language institute's headquarters in Taipei, Taiwan. The parents at the time of data collection will have one or two children enrolled at the language institute.

To qualify for participation, teachers should have worked in one of the language schools for more than six months. School managers should have worked as managers for at least one year. Parents will have to speak some English

and be willing to be interviewed; and the curriculum writer will have to have at least two years of curriculum writing experience. The participants will be identified on the basis of availability and willingness to participate in the interviews.

Materials

Some classes will be observed and field notes will be made with the use of an observation template. Four sets of interview questions were designed to elicit responses from the four groups of participants. These are mostly open-ended questions combined with some specific questions relating to the participants' views on literacy and teaching literacy skills (i.e., reading and writing). The interviews will provide me with valuable insights and a deeper understanding of the participants in their context.

Two different sets of curriculum materials will be reviewed. The first set of materials will consist of reading books, writing books, and teacher manuals that are used to prepare learners for elementary school. Additional curriculum materials that will be considered are the books used in the first semester of elementary school. These materials are not associated with the Taiwanese elementary school system but are part of the bushiban system; that is, they are English class materials not issued by elementary schools but by the educational institute. These materials consist of a textbook, a workbook, and two homework books.

Procedure

Interviews will take place in a variety of locations, such as the school itself, the head office, coffee shops, and the homes of the Taiwanese parents. All of these locations are in Taipei city in the Da-An, Sinyi, and Songshan districts. The exact choice of location for the interviews is

left to the participants, and the interviewer/researcher will accommodate those requests.

The classroom observations will be done at two different schools. No recording of classes is allowed. In all five observations, I will observe the class through the visitor's window. These classrooms typically have three solid walls and one wall that includes a large window directed to the inside of the school. This is often referred to as the parent window or visitor's window. Through the visitor's window, all classroom activities can be seen and heard. The focus of the classroom observations is threefold. I will take note of the modality (i.e., written or oral) that the young learners practice, the time spend on that modality, and the types of activities.

Two possible risks are considered. First, language difficulties could be a problem when interviewing the Chinese school managers and parents. I will allow participants to answer in Chinese if they are not sure about the English vocabulary. I have limited Chinese skills and am willing to use the services of an interpreter. I will also attempt to interview parents with a high intermediate English ability to lessen ambiguity during the interviews. Second, the teachers could experience discomfort in criticizing the curriculum or in being observed. Because I am a former teacher and fully aware of possible conflicts, I will assure the teachers that their views will be kept confidential.

One of the caveats of this kind of research is that the presence of the observer might cause the participants to act differently. To prevent my observation from influencing the linguistic behavior of those being observed, I will attempt to enhance

credibility by collecting data over a period of eight weeks to ensure that the participants have become used to me and are behaving naturally.

Analysis

As is typical in qualitative research, the data will be analyzed through an inductive approach in which themes and patterns emerge from the data. All interviews will be transcribed, and these transcriptions will be entered into NVivo 8. The data will be read, and a list of general themes will be compiled.

References

- Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *Modern Language Journal*, 93, 227–240.
- Buckwalter, J. K., & Lo, G. H. (2002). Emergent biliteracy in Chinese and English. *Journal of Second Language Writing*, 11, 269–293.
- Butler, Y. G. (2004). What level of English proficiency do elementary school teachers need to attain in order to teach EFL?: Case studies from Korea, Taiwan, and Japan. *TESOL Quarterly*, 38, 245–278.
- Butler, Y. G. (2005). Comparative perspectives towards communicative activities among elementary school teachers in South Korea, Japan, and Taiwan. *Language Teaching Research*, 9, 423–446.
- Butler, Y. G. (2005). English in elementary schools: Current English language education policies in South Korea, Japan, and Taiwan. *The Journal of Teaching for Young Learners of English*, 1, 49–73.
- Butler, Y. G. (2007). How are non-native English speaking teachers perceived by young learners? *TESOL Quarterly*, 41, 731–755.
- Chien, S. C. (2011). Discourse organization in high school students' writing and their teachers' writing instruction: The case of Taiwan. *Foreign Language Annals*, 44, 417–435.
- El-Koumy, A. S. A. (1998). Effect of dialogue journal writing on EFL students' speaking skill. Retrieved from ERIC database. (ED 424772)
- Elley, W. B., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, 19, 53–67.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59–84.
- Gorsuch, G. J. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34, 675–710.
- Gunderson, L. (2009). *ESL (ELL) literacy instruction: A guidebook to theory and practice* (2nd ed.). New York, NY: Routledge.
- Hardison, D. M. (2011, March). *L2 engagement, perceived competence, and communication affect: Predictors of competence in the component skills of L2 oral interactions*. Paper presented at the meeting of the American Association for Applied Linguistics Conference, Chicago, IL.
- Harklau, L. (2000). From the "good kids" to the "worst": Representations of English language learners across educational settings. *TESOL Quarterly*, 34, 35–67.
- Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, 11, 329–350.
- Kim, Y. (2008). The effects of integrated language-based instruction in elementary ESL learning. *Modern Language Journal*, 92, 432–451.

- Lai, Y. C. (2009). Language learning strategy use and English proficiency of university freshmen in Taiwan. *TESOL Quarterly*, 43, 255–280.
- Leki, I. (2000). Writing, literacy, and applied linguistics. *Annual Review of Applied Linguistics*, 20, 99–115.
- Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology*, 39, 1005–1019.
- Matsuda, P. K., & De Pew, K. (2002). Early second language writing: An introduction. *Journal of Second Language Writing*, 11, 261–268.
- Maxim, H. H. (2002). A study into the feasibility and effects of reading extended authentic discourse in the beginning German language classroom. *The Modern Language Journal*, 86, 20–34.
- Nassaji, H., & Cumming, A. (2000). What's in a ZPD? A case study of a young ESL student and teacher interacting through dialogue journals. *Language Teaching Research*, 4, 95–121.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45.
- Polio, C. (2012). The acquisition of second language writing. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 319–334). New York, NY: Routledge.
- Rubin, D. L., & Kang, O. (2008). Writing to speak: What goes on across the two-way street. In Belcher, D. & Hirvela, A. (Eds.), *The oral/literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 210–225). Ann Arbor, MI: University of Michigan Press.
- Strube, S. (2011). Conveying meaning: Oral skills of L2 literacy students. *TESOL Quarterly*, 44, 628–637.
- Tarone, E., Bigelow, M., & Hansen, K. (n.d.). *The impact of alphabetic print literacy level on oral second language acquisition*. Retrieved from <http://www.theliteracyinstitute.org/projects/pdf/LESLLA%20Research6.pdf>
- Valdés, G. (1992). Bilingual minorities and language issues in writing: Toward profession-wide responses to a new challenge. *Written Communication*, 9, 85–136.
- Williams, J. (2008). The speaking-writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral/literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 10–25). Ann Arbor, MI: University of Michigan Press.
- Weber, R., & Longhi-Chirilin, T. (2001). Beginning in English: The growth of linguistic and literate abilities in Spanish-speaking first graders. *Reading Research and Instruction*, 41, 19–49.
- Weissberg, R. (2006). Developmental relationships in the acquisition of English syntax: writing vs. speech. *Learning and Instruction*, 10, 37–53.
- Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, 23, 345–368.
- Yang, N. D. (1999). The relationship between EFL learners' beliefs and learning strategy use. *System*, 27, 515–535.
- Yu, M. C. (2008). Teaching and learning sociolinguistics skills in University EFL classes in Taiwan. *TESOL Quarterly*, 42, 31–52.

ESL Reading Test Development and Analysis

Hyojung Lim

Michigan State University

hyojung@msu.edu

Young-Shin Kwon

Teachers College, Columbia University

yshkwon79@gmail.com

The purpose of this paper is to present an overview of ESL reading test development and analysis in the context of a small-scale ESL classroom. We created 12 multiple-choice items for an ESL reading mid-term exam, administered the test in the Community English Program at Teachers College in fall 2008, and analyzed the test results to evaluate the reliability and validity of the test. We first describe the nature of reading ability by reviewing the literature on second language reading and reading assessment. Based on a widely shared definition of reading ability, we suggest a theoretical construct of reading ability and relevant observable variables. Following Bachman and Palmer (1996), this paper provides practical guidance for language teachers with regard to how to create reading test items and assess the test quality from describing the target language use (TLU) domain and task types, developing a test design statement, generating the blueprint for test operationalization, coding multiple choice items, to conducting item and distractor analyses. Issues revolving around L2 reading test development are further discussed.

We created this test as part of the mid-term exam designed for the ESL learners of the Advanced 2 (A2) class of the Community English Program (CEP) at Teachers College, Columbia University in New York City. The CEP is a lab-school where Teachers College students from the Applied Linguistics and TESOL program teach adults from the community as part of their practical training, apply various teaching methods based on linguistic and pedagogical theories, and collect data for empirical studies related to the instruction and assessment of second-language (L2) learners. The students of CEP are adult ESL learners, most of whom are either immigrants, international students who are planning to study or already in school, or family members of international students in the Columbia University community.

The CEP curriculum consists of twelve levels in total, ranging from basic (B1 to B4), intermediate (I1 to I4), and to advanced (A1 to A4).

Participants of this study are advanced ESL students in the A2 evening class, who are still aiming for higher-levels of English proficiency, some seeking to advance to the levels of A3 and A4. About 67% of the students completed either graduate or post-graduate degrees and 50% are planning to stay in the United States for an academic or a vocational purpose. The A2 course focuses on further developing the four integrated skills in English. The course objectives are to improve students' skills with a focus on critical analysis and self-expression, and to help students with their knowledge and application of pragmatics. The mid-term exam accounts for 30% of their final

grade. By the time of data collection, the class theme was about the political or social issues as dealt with in the 2008 US presidential election. With regard to the language skills, students were taught lessons on (1) how to use contextual clues, analyze arguments, make inferences and generalizations and determine the purpose and function of a text for reading, (2) how to link paragraphs to essays, create an argument, organize information and use transitions for writing, (3) how to summarize, identify implications, and personalize the information for listening, and (4) how to continue a discussion, present ideas and debate a topic for speaking. In terms of reading, students were exposed to extensive reading of news articles both in and out of class.

The mid-term exam we designed for the A2 class can be classified as an achievement test or a progress test, given that it aims to measure the extent of learning or mastery within a specific instruction domain. The test result, as part of final grade, is used to make decisions about their advancement or competency. The mid-term may serve as a diagnostic test as well: the test result carries information about students' strengths and weaknesses, and thus can prescribe future teaching or learning directions for the rest of *SEM*ester. Since the purpose of the course is to improve integrated skills, listening and speaking are supposed to be assessed in the test. However, due to time constraints as well as test practicality, the mid-term includes only grammar, listening, reading and writing. In this paper, we focus on the reading test, since reading skills were more emphasized in the class during the first half of the semester than listening and speaking skills.

We will first describe the nature of reading ability based on the review of prior research on second language reading and assessment. Based on the prior literature on reading comprehension, we suggest a theoretical construct of L2 reading ability. The theoretical construct of reading ability provides a useful ground for the subsequent test construction: describing the target language use (TLU) domain and task types, writing test design statements, developing the blueprint for the test operationalization, coding multiple choice section, and finally administering the test. Lastly, the test reliability and the construct validity will be assessed through item analyses.

Reading Ability

To measure learners' reading ability in the A2 class at CEP, essential is to first clarify what reading ability is and/or what reading components the test is to assess. Reading is a complex, multifaceted cognitive behavior that involves a number of linguistic and cognitive processes. Thus, it seems hardly possible to come up with one simple definition for it (Grabe & Stoller, 2002). Instead, many reading researchers have shed light on multiple aspects of the reading construct. Researchers have foraged for discrete factors that constitute L1 and L2 reading comprehension (Barnett, 1986; Devine, 1981), identified cognitive processes involved in different types of reading (Weir, Hawkey, Green, & Devi, 2009; Khalifa & Weir, 2009), and investigated strategies/skills that learners likely employ while reading L2 texts (Cohen & Upton, 2007; Savery, 2012; Sheorey & Mokhtari, 2001).

According to the information-processing approach, reading comprehension is considered as the

product of bottom-up and top-down reading skills. Grabe and Stoller (2002) characterize reading as a serial process consisting of two different levels: lower-level and higher-level processes. Lower-level processes include basic linguistic processes such as word recognition, syntactic parsing, and even simple sentence verification. Reading begins with decoding a string of letters in print, recognizing word meanings, parsing sentence structures, and finally to constructing clause-level, textual meaning units. To obtain a high level of comprehension, therefore, it is crucial for learners to be able to execute the lower-level processes automatically. Efficient processing frees up available mental resources, which eventually helps readers to hold more information in their memory (Daneman & Carpenter, 1980). Both L1 and L2 reading researchers have acknowledged the contribution of automatic bottom-up processing skills to the increased reading comprehension (Koda, 2005; Roberts, Christo, & Shefelbine, 2011). Conceivably, without processing lexical and syntactic information, readers cannot run any higher-level cognitive processes (e.g., inferences) where we believe ultimate comprehension takes place. In L2 reading, Alderson (1984) claims that foreign-language reading is a language problem rather than a reading problem; especially for those who are already literate in their L1, much of the difficulty in L2 reading comprehension could be mainly due to their language proficiency, not to their literacy skills. This is particularly true for educated adult language learners who already possess higher-order thinking ability in their native language but lack automatic processing skills in the L2. In terms of assessment, any reading tests are likely to assess lower-level linguistic processes

in an implicit way; there is no reading test item that directly measures test takers' word recognition skills or sentence processing skills. Instead, bottom-up skills are often assumed to be tested in a rather unified or general way (Alderson, 2000).

The top-down approach to reading underscores the effects of higher-level reading processes on comprehension. This is where the schema theory comes into play. According to Grabe and Stoller (2002), the higher-level processes begin to play a role in the text model of comprehension, where readers draw main ideas and supporting details from a text at or beyond the clause-level meaning units. While reading, readers are likely to activate their content and formal schemata: content schemata means readers' background knowledge of the content area of the text, whereas formal schemata pertains to readers' knowledge of the rhetorical structures of different types of texts (Carrell & Eisterhold, 1983). The essential idea of the schema theory is that readers' familiarity with the discourse organization as well as with the topic facilitates their understanding of the text. Thus, reader variables, such as cultural background or topical knowledge, often become determining factors for the quality of comprehension. Finally, Grabe and Stoller explain that executive control (or metacognitive) processes are part of the higher-level reading processes. Previous empirical studies found that good readers have advanced synthesis and evaluation skills so that they can simultaneously monitor their comprehension and quickly adopt relevant reading strategies (Paris & Myers, 1981). In the context of L2 reading assessment that measures both language and reading ability, however, we believe that educated adult L2

learners should be forced to utilize their L2 linguistic knowledge and skills rather than their content knowledge, or general reasoning ability. Especially, international graduate students who usually have high-level literacy skills in their L1, meaning that they know how to approach a text and how to inspect their own understanding. As long as they meet the threshold of L2 language proficiency, if any, such learners should be able to transfer their cognitive and literacy skills to the second language (Cummins, 1991).

From a balanced perspective, the interactive model highlights that the bottom-up processing works in concert with the top-down processing, or vice versa. Interaction has been understood in many different ways. The “simple view of reading” proposed by Hoover and Gough (1990) views reading comprehension as the combination of word decoding and listening comprehension; lacking either decoding skills or listening ability can deteriorate the quality of reading comprehension. Rather, taking a “compensatory” approach, Stanovich (2000) points out the tendency of readers resorting to their higher-level processing skills to compensate for their deficiency in lower-level processing skills. For instance, readers often use context clues to guess the meaning of an unknown word and consequently improve their understanding of the text. In L2 reading, Bernhardt's (2005) compensatory model echoes Stanovich's view, thereby describing how L2 readers rely on their L1 literacy skills to improve L2 language-processing skills or how an increase in word knowledge helps to accelerate the processing of L2 sentences. Meanwhile, Grabe (1991) suggests a more general type of interaction; the interaction between a

text and a reader. Readers form their reading comprehension by relating the given textual information to their background knowledge. Given that it is a reader who reconstructs the representation of a text, the way that the reader processes the text likely determines the type and level of comprehension. To us, the interaction discussed in Grabe seems to rather support the schema theory where high-level reading processes play a substantial role.

Another way to approximate the reading construct is to explore types of strategies that readers employ while reading. Researchers, in their examination of good and poor readers, have discovered that good readers are likely to adopt various effective reading strategies (Anderson, 1991; Ebrahimi, 2012; Paris, Limpson, & Wixson, 1983; Paris & Myers, 1981). In this regard, Grabe (2004) states that “a number of individual comprehension strategies have been shown to have a significant impact on reading comprehension abilities” (p.51). According to Fitzgerald (1995), reading strategies can be understood in two different ways: (a) psycholinguistic strategies that learners use to recognize and comprehend lexical items; and (b) metacognitive strategies that learners use to deal with a whole text and repair miscomprehension. The psycholinguistic strategies are similar to the compensatory strategies that L2 learners rest on to overcome linguistic limitations. In the context of assessment, Cohen and Upton (2007) documented the reading strategies based on international students' verbal reports. The observed strategies were categorized into three groups: (a) approaches to reading the passage (e.g., considering prior knowledge of the topic), (b) uses of the passage and the

main ideas to improve understanding (e.g., re-reading to clarify the ideas), and (c) identification of important information and the discourse structure of the passage (e.g., looking for sentences that convey the main ideas). Note that the reading strategies listed here are all language-independent, metacognitive strategies. According to Fitzgerald's (1995) collection of literatures on L2 reading strategies, the most common were: asking questions, rereading, imaging, using a dictionary, anticipating or predicting, reading fast or changing speed, associating, skipping, and summarizing. From learners' perspective, Judith (1995) discovered that scanning for specific information, skimming, re-reading, word-guessing skills and summarizing were valued most by students learning Spanish as a second language. Taken together, L2 readers use various types of strategies at all levels (e.g., lexical, sentential, and textual level) to maximize their comprehension. They are likely to approach L2 reading as a problem-solving task, thereby evoking higher-order cognitive processes (e.g., monitoring), presumably in the same way that they would do in L1 reading. The reading strategies reviewed so far can be reduced to three major reading behaviors: reading to search for information, integrating pieces of information, and figuring out hidden meanings (e.g., an author's intention).

Lastly, but most importantly, the purposes of reading need to be taken into consideration, as reading itself is a purposeful behavior. According to Carver (1997), there are two types of reading: "reading" and "reading to learn." The term "reading" pertains to basic comprehension — reading a text to understand major points — while "reading to learn" involves the

reconstruction of a text — figuring out main ideas and supporting details. Similarly, drawing from the cognitive processing model for reading comprehension, Khalifa and Weir (2009) propose two kinds of reading at two different levels: careful and expeditious reading at the local and global level, respectively. Careful reading is intended to extract complete meaning from a given text (Hoover & Tunmer, 1993). It is conceived as slow, careful, linear, and incremental reading. Conversely, expeditious reading is rapid, selective, and efficient reading, including scanning and skimming. Both readings can take place at the lexical or sentential (local) level, or at the paragraph or textual (global) level. Albeit using different terms, ETS (2000) suggests the purpose-driven framework for the iBT TOEFL reading test: reading to find information, reading for basic comprehension, reading to learn, and reading to integrate information. In light of item difficulty, reading to integrate information is thought to be more difficult than reading to find information, since the former requires relatively higher-order cognitive abilities. Taking learners' proficiency into account, we decided to include more inference-type questions. For advanced learners, such as those in the A2 class, reading should not be a language problem any longer. Rather, they are expected to read to synthesize and critique texts.

To sum up, the reading construct that we want to measure in the mid-term exam entails three variables: gist, details, and inference. The information-processing perspective on reading, the skill-and-strategy approach, and the reading-purpose perspective all provide strong rationale for the variables that we suggest. Given that lower-level processes are assessed in an implicit manner, we

expect our students to be able to make connections across sentences and paragraphs quickly and accurately so that they can correctly comprehend main ideas and supporting details in a given time. In terms of reading types to be tested, reading for gist and details may be associated with search reading, skimming, and reading to learn. More specifically, reading for gist can be involved in such items as summarizing a text, finding a main idea, or selecting a headline/title for the text. Reading for details can be induced by the items such as finding specific information, relating a pronoun to its referent, and rephrasing a given sentence. For the inference questions, learners have to make use of their content and formal schemata to answer the questions. Readers will be asked to derive both literal and implied meaning at lexical, sentential, and textual level, to guess an author's intention for using specific expressions in the flow of ideas, and to read an author's tone. Figure 1 summarizes the theoretical model of reading ability for the reading test in CEP A2 mid-term test.

Test Construction

The Target Language Use (TLU) Domain

The context of the target language use (TLU) domain is the CEP evening class of L2 learners at the Advanced 2 level, taught in a classroom at Teachers College. The class integrates all four language skills of reading, writing, listening, speaking, and includes grammatical contents, while following a weekly theme-based curriculum. The learners are all adults from various nationalities and cultural backgrounds, coming from different occupational backgrounds as well. They are generally enrolled in the CEP to advance their English proficiency overall, while some learners have specific purposes such as to enter a higher-education institution or an English-speaking workplace in the United States. While it would be difficult to pinpoint a specific TLU domain because of the broad background of the group of learners, we have decided that language instruction would be the most appropriate TLU domain for our subjects.

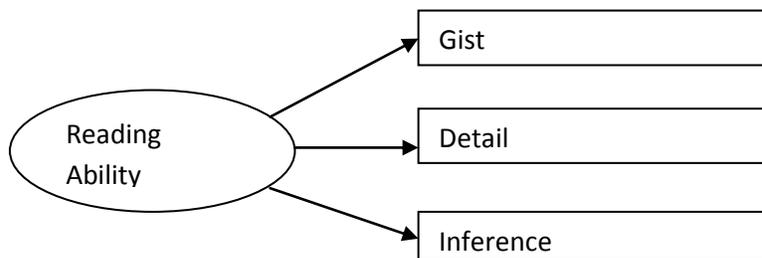


Figure 1. A theoretical model of reading ability.

We attended one class session to observe what themes the learners were specifically dealing with after looking over the syllabus and textbook for the class. At the time of observation, the class topic was the current presidential candidates and their campaigns on several socio-economic issues. The learners had been assigned to research information from the media and support one of the presidential candidates' campaigns on a specific issue (e.g., education, health care, and energy) based on their investigation. In pairs, they were to present a brief spoken debate on their ideas by supporting them with the information they found. According to the syllabus, the learners had also been instructed on writing academic essays on opinion-based subjects. We judged that after completing these instructional tasks, the learners should be able to understand and analyze fact-based information from the media. They should be able to read for specific details and infer further facts according to the given information. In both academia and the workplace, extracting information from the media or other informational sources, and making critical judgments of the information to form individual opinions are important abilities that are often required to competently perform a given duty, such as making decisions about a course of action or a direction that a business should take.

Taking these into consideration, for the reading task of our test we decided to use a news article on the subject of education and the differing views that the presidential candidates have on this issue. The skills needed for the reading tasks are (a) reading for gist (both at the passage level and paragraph level), (b) reading for detail such as for finding facts and correct word references, and

(c) making correct inferences about the writer's purpose or rhetorical purpose.

Design Statement

Following Bachman and Palmer (1996), we developed a design statement for the current test (see Table 1). The design statement is essential for the subsequent procedure of test development, operationalization, trialing, and assessment use. Based on the design statement, the test structure and the task specifications are presented in the following section.

Operationalization

Test structure.

1. Number of tasks: The test consists of one task containing 12 items to measure the test takers' ability of reading in a language-instructional domain. The students must read a news article and answer twelve multiple-choice questions.

2. Salience of tasks: The reading task is clear, with clear labels and specific instructions provided.

3. Relative importance of tasks: All items within the task are of equal importance and worth the same amount of points in the mid-term exam.

4. Number of tasks per part: The reading part is one task consisting of a set of twelve multiple-choice questions. Table 2 provides a summary of the test structure.

Table 1
Design Statement

1. Test purposes	
A. Inferences B. Decisions I. Stakes II. Individuals affected III. Specific decisions to be made	<p>About test-takers' reading and writing ability in a language-instructional domain.</p> <p>Relatively high in the context of the course; results are used to determine advancement to the next level in the program (the mid-term examination counts for 30% of the final course grade).</p> <p>Test-takers (CEP students) and CEP class teacher</p> <p>1. Achievement</p> <p>a. Progress: To determine if students have mastered the language skills covered up to the mid-term exam.</p> <p>b. Grading: Results are part of the mid-term grade, a component of the final grade, which determines advancement to the next level in the program.</p> <p>2. Diagnosis</p> <p>a. For teachers: To evaluate each student's strengths and weaknesses in order to help students make further improvement.</p> <p>b. For students: To obtain information on their own strengths and weaknesses in order to identify and overcome weaknesses.</p>
2. Description of TLU domain and task types	
A. Identification of tasks 1. TLU domain 2. Identification and selection of TLU tasks for consideration as test tasks B. Description of TLU task types	<p>Language-instructional, but also possible to be real-life for some students.</p> <p>TLU tasks to be analyzed were identified based on the course syllabus and class handouts. The Reading Task (reading a news article and answering multiple-choice questions) is an instructional task similar to those performed in class. Reading a news article can also be a real-life task.</p> <p>Refer to Table 3 for the test task specifications.</p>
3. Definition of constructs	
A. Language ability	<p>The construct definition for this achievement test is based on both a theoretical model of language ability and the content of the class. The elements of language knowledge included in the construct definition are:</p> <ul style="list-style-type: none"> - Reading ability <ul style="list-style-type: none"> a. Reading for gist (summary, main idea, title) b. Reading for details (fact finding, word reference)

B. Strategic competence	c. Inferencing (writer's purpose, rhetorical purpose, tone) Not included in the construct.
C. Topical knowledge	Not included in the construct. However, some degree of topic knowledge is assumed, as students are familiar with the topic dealt with in class (e.g., education, presidential candidate's debate).

Table 2
Test Structure

Construct	Task Type	Number of Tasks	Number of Items	Time	Scoring
Reading Ability • Gist • Detail • Inference Theme: Education issues in the US presidential debate	Selected-Response (Multiple Choice)	1	12	30 mins	Dichotomous Scoring 0/1 12 points available

Test task specifications.

- 1) Purpose: See the design statement (Table 1)
- 2) Definition of construct: See the design statement (Table 1)
- 3) Setting
 - a) Physical characteristics: Classrooms (Horace Mann Hall Rm. 136) at Teachers College, Columbia University. See the test task specifications in Table 3.3 for a detailed description of classroom conditions.
 - b) Participants: The CEP teacher and the test-takers (CEP students).
 - c) Time of task: During class hours on Thursday, October 23, 2008.
- 4) Time allotment: Thirty minutes.
- 5) Instructions:
 - a) Language: The target language (English) because test-takers have a variety of native languages. Separate instructions are provided for the reading part and the writing part, and the students are allowed to ask questions about instructions they are not sure of.
 - b) Channel: Visual (writing).
 - c) Instructions: See the copy of the test provided in Appendix C
- 6) Characteristics of input and expected response: See the test task specifications in TABLE 3 (Appendix A).
- 7) Scoring method:
 - a) Criteria for correctness: The multiple-choice questions are scored dichotomously based on an objective answer key.
 - b) Procedures for scoring the responses: The multiple-choice questions are scored dichotomously based on the objective answer key for the multiple-choice questions. One point is given for each correct answer and zero points are given for each incorrect answer, for a possible total of twelve points.

- c) Explicitness of criteria and procedures: The test-takers are informed in general terms about the scoring criteria in the instructions. Table 3 summarizes the task specifications for each task. (See Appendix A)

Item Coding

The reading test consists of twelve multiple-choice items that are divided into three observable variables: understanding the gist, finding details, and making inferences. An inference is an overarching notion of guessing from the context, ranging from guessing meanings of new words to reading the author's tone. Table 4 illustrates the observable variables for each item and a brief description of their subordinate variables.

Administration Procedures

The test was administered to the CEP Level A2 evening class as part of their mid-term evaluation and took place in their original classroom during their usual class time. The students were given a separate listening and grammar test at the beginning (given by the CEP instructor) and afterwards handed out the test booklets on the reading and writing parts, which they could start immediately upon receiving it. As the test booklets were being handed out, the students were told how much time they had to complete the test, and were allowed to leave the classroom upon the completion of the test. The students were allowed to ask any questions that arose while taking the test. The entire test period lasted for two and a half

Table 4
Coding Multiple Choice Items for Reading

Observed Variable	Item Number	Description of the item
Gist	1	Giving a title to the entire reading passage.
	11	Understanding the main idea of a paragraph.
	12	Summarize the entire passage.
Inference	2	Reading the author's tone.
	3	Understanding a rhetorical purpose.
	4	Guessing an expression in context.
	6	Understanding a rhetorical purpose
Detail	7	Understanding the metaphoric use of a word.
	5	Comprehending specific information explicitly stated in the text.
	8	Comprehending specific information explicitly stated in the text.
	9	Finding the referent of a pronoun.
	10	Rephrasing a sentence.

hours, which included all four parts of the test (listening, grammar, reading, and writing). As the test-takers finished their test, they were given a brief post-test survey to fill in. The survey consisted of general-information questions about the test-taker (age, nationality, occupation, etc.) and some questions about self-perceived language proficiency and qualities of the test itself.

Test Takers

The number of the students was twelve from the evening A2 class of Community Language program at Teachers College in New York. Most of them were in their late twenties or early thirties, while one was in her late thirties and one in her early forties. Ten students completed their education at or beyond the graduate level, while only two students obtained up to a bachelor's degree. Their majors were as diverse as Social Work, Trading, Economics, Law Administration, English and American Literature, Electronic Engineering, Art Design, and Home Economics. Regarding nationality, the East-Asian students were dominant: seven from Japan, two from Korea and one from China. The two remaining participants were from Bolivia and Poland. Nine students were female and three were male. With regard to the length of stay in the United States, it varied from one and a half months to eight years. To be more specific, eight students (67%) lived in America for less than one year, while two students for more than five years. The post-test questionnaire was used to collect the participant information (Appendix B).

Test Instrument

The purpose of our test was to measure reading ability within a specific instructional domain. By reviewing various articles on reading ability, we have decided to include gist, inference and detail for the reading construct. The reading test consisted of twelve multiple-choice items; 3 for gist, 5 for inference and 4 for detail variable. The topic of the reading task was "American education" discussed in the 2008 US presidential election. A copy of the actual and the expected responses are attached in Appendix C.

Scoring Procedures

The multiple-choice reading task was scored objectively and dichotomously. One scorer rated every test paper using an objective answer key and assigned one point to correct answers and zero points for incorrect answers. The total score was the sum of the point that each item earned. The possible range of scores on this task was therefore 0 to 12.

Analyses and Results

Descriptive Statistics

The reading section had 12 multiple-choice questions, for a total possible score of 12 ($k=12$). One point was assigned to a correct answer and zero to an incorrect answer. In terms of the measures of central tendency, the mean was 6.75 (56.25%), the median was 6.50, and the mode was 6.00. The skewness value of the score distribution was -0.04 . The kurtosis was 0.34. The kurtosis indicates the degree to which the distribution is peaked. Given that the skewness value and the kurtosis were close to zero, the test scores were normally distributed. In terms of the data dispersion, the range was 7.00, from a minimum score of 3.00 to a

maximum score of 10.00. The standard deviation was 1.91. The results are summarized in Table 5.

Table 5
Descriptive Statistics for the Reading Task

Statistics	Results
Number of participants (N)	12.00
Number of items (k)	12.00
Maximum possible score	12.00
Mean	6.75
Median	6.50
Mode	6.00
Skewness	-0.04
Kurtosis	0.34
Range	7.00
Minimum	3.00
Maximum	10.00
Standard deviation	1.91

Considering that the test was an achievement test for the A2 class at CEP, we expected the distribution of scores to be negatively skewed, and ideally students were to answer 70% of the test correctly on average. However, our test results turned out to be undesirable for a criterion-referenced test: the skewness value of -0.041 and the kurtosis of 0.334 indicate that the test scores were normally distributed. Furthermore, the average of 6.75 means that only 56.25% of the test was answered correctly on average, which was somewhat lower than the cut-off line (70%) for the pass and fail standard at CEP.

From the statistical figures, we could infer that our test was somewhat difficult for the participants. Presumably, only a few students might have mastered the theme and the reading strategies previously taught in class. The larger proportion of inference questions might

have raised the level of difficulty in that these questions usually require higher-order cognitive skills. Therefore, it could be that our test failed to correctly measure students' reading ability on the basis of the class objectives.

The standard deviation of 1.91, the kurtosis of 0.34 and the range 7.00 out of 12.00 suggest that the test scores are somewhat widely spread out. Thus, the group in the evening A2 class proved to be heterogeneous with regard to English reading ability. It may be that these students had not been correctly placed in the beginning, or has truly shown varying degrees of development in reading comprehension. In Table 6, the results are illustrated in the stem-and-leaf plot.

Table 6
Reading MC Stem-and-Leaf Plot

Frequency	Stem	Leaf
1	3.	0
1	5.	0
4	6.	0 0 0 0
3	7.	0 0 0
2	9.	0 0
1	10.	0

Internal Consistency Reliability and Standard Error of Measurement for the MC Task

This section evaluates the test reliability. Test reliability means the extent to which the results are consistent or stable. To be more specific, the reliability estimates are interpreted as the percent of systematic, consistent, or reliable variance in the scores of a test, including both true and random error variance. When it comes to the MC items, the internal consistency reliability across the 12 items was examined by calculating the reliability coefficient. The internal-consistency reliability informs us as to the degree to which each item

relates to all the other items. Subsequently, we calculated the standard error of measurement (*SEM*) to determine a confident interval of a student's score; the narrower *SEM* evidences the higher test reliability, meaning that test cores will less fluctuate if the test is repeated. We also calculated Cronbach's alpha as an alternative measure of the split-half reliability. The split-half reliability was not appropriate for this short test, because the number of test items was too small to separately score and compare the odd-numbered and the even-numbered items. Table 7 presents the internal consistency for the 12 MC items.

Table 7
*Internal Consistency Reliability
Statistic for the Reading Task (K=12)*

Cronbach Alpha Coefficient	Number of Items
0.343	12

Cronbach's alpha typically ranges from 0 to 1, with 1 being the most consistent. The coefficient 0.343 suggests that the internal consistency for our MC items were relatively low. With the reliability of 0.343, the scores are around 34% consistent. That leaves 66% of measurement error or random variance in the scores. This implies that the degree to which the items relate to one another was somewhat low, so was the internal consistency of the test. There are several reasons for the unexpected results: First, the small number of items might be ascribed to the low consistency. The MC items were only twelve in total, consisting of three items for the gist, four for the detail, and five for the inference variable. Hence, every correlation between items should have a substantial impact on the

reliability of the test. Second, the sample size of twelve students might have been too small to correctly calculate the reliability coefficient. Only a couple of students' mistakes in their responses could have affected the statistical analyses. In either case, the low internal consistency seems mainly due to the limited amount of data. All in all, we do not have sufficient evidence to say that our test is trustworthy. In addition to Cronbach's alpha, the *SEM* was calculated to determine the band around a student's score within which the student's score would probably fall, if the test were repeated. This gives an idea of how accurate an individual's true test score might be. The computation formula for *SEM* is given in Table 8, where the result for our test is summarized as well.

Table 8
*Standard Error of Measurement for the
Reading Task*

$$*SEM = S \sqrt{1 - r_{xx}}$$

$$SEM = 1.913 \times 0.811 = 1.551$$

* where S = standard deviation (retrieved from the descriptive statistics) and r_{xx} = reliability estimate for the test, which is equal to the Cronbach's alpha coefficient

Based on the estimated $SEM = 1.551$, a 95% confidence ($\pm 2 SEMs$) interval was calculated. According to the result, a student's score would consistently fall within a band of two *SEMs* higher and two *SEMs* lower than her raw score 95% of the time if s/he were to take the test multiple times. For instance, participant #3 received 7 out of 12, but it is 95% certain that the score would fall somewhere between 3.174 and 10.102 if the participant were to take the same test repeatedly. Since each item was scored dichotomously, we rounded up these values to 4 and 11, respectively.

Considering that the total reading score was 12, the *SEM* of 1.91 seems relatively large for the short test, and thus the 95% confidence interval for participant #3's score turned out to be too broad. This indicates that extra factors, other than one's reading ability, may have confounded the observed scores such as the degree of motivation, fatigue, and chance knowledge of item content.

Item Analysis

To search the causes for the low internal consistency, the 12 MC items were analyzed by calculating the item difficulty (or *p*-value), the item discrimination index (or *d*-value) and the "alpha if item deleted." To explain each term briefly, the item difficulty is an index that tells us the proportion of test takers who got the item correct in proportion to all the test takers who answered the item. The item discrimination indicates the degree to which the item discriminates between different groups. By convention, the high 27% of the students is compared with the low 27% in a norm-referenced test. Lastly, the "alpha if item deleted"

shows a recalculated Cronbach's alpha if the item is deleted from the test. These statistical results were the bases for the decisions made on whether to delete or keep each item (see Table 9).

The *p*-values ranged from 0.167 (for item 6) to 0.917 (for item 11). In other words, item 6 was extremely difficult, therefore, only two participants got the answer correct, while item 11 was extremely easy, therefore, everyone except for one participant got it correct. The overall *p*-value of the twelve items was 0.576.

Given that an ideal achievement test aims for a *p*-value of 0.70, our test appeared to be somewhat difficult as a criterion-referenced test, which is consistent with the earlier report on the descriptive statistics. Except for items 8, 9 and 11 with *p*-values of 0.833, 0.833 and 0.917 respectively, the *p*-values of all the other items were lower than 0.70. Moreover, item 5, 6 and 10 were extremely difficult with *p*-values of 0.333, 0.167 and 0.250 respectively. By only looking at the estimated *p*-values,

Table 9
Item Analysis for the Reading Test

Item	Observed Variable	Difficulty (<i>p</i> -value)	Discrimination (<i>d</i> -value)	Alpha if item deleted	Decision
1	Gist	0.417	0.239	0.269	Keep
2	Inference	0.417	0.133	0.317	Keep
3	Inference	0.583	0.412	0.185	Keep
4	Inference	0.667	-0.417	0.516	Delete
5	Detail	0.333	-0.249	0.461	Delete
6	Inference	0.167	0.106	0.328	Keep
7	Inference	0.667	0.367	0.328	Keep
8	Detail	0.833	0.240	0.284	Keep
9	Detail	0.833	-0.021	0.367	Not sure
10	Detail	0.250	0.706	0.068	Keep
11	Gist	0.917	-0.189	0.396	Not sure
12	Gist	0.667	0.249	0.267	Keep

our test seems more like a placement or a proficiency test, rather than an achievement test. In terms of the difficulty level of each variable, the average p -value for gist items was 0.667, that of inference items 0.500, and that of detail items 0.562; inference items were relatively more difficult than the other two variables, as we expected.

To calculate the discrimination index, the point biserial correlation was utilized. The “corrected item-total correlation” was interpreted as the d -value. By convention, the items with a d -value of 0.40 and above are evaluated as very good items. Those with a d -value of 0.30 to 0.39 are considered as reasonably good items, but subject to improvement. On the other hand, items with a d -value of 0.20 to 0.29 do not effectively differentiate the high 27% from the low 27% of test-takers. Lastly, a d -value of 0.19 and below indicates that the item needs to be deleted or improved. Based on this standard, only three items (item 3, 5 and 10) were evaluated as the very good or relatively good items with the d -value of 0.412, 0.367 and 0.706, respectively. Five items were evaluated as either marginal (item 1, 8, and 12) or poor items (item 2 and 6) and thus presumably need to be deleted or revised. Lastly, four items (item 4, 5, 9 and 11) were almost non-discriminating or negatively discriminating with the d -value of -0.417 , -0.249 , -0.021 and -0.189 respectively. Overall, nine out of twelve items were labeled as marginal, poor and negatively discriminating items due to their low or negative d -values. Our conjecture is that the test might have been simply too difficult for all students. Both the high-scoring and the low-scoring group seem to have missed the same questions. Another possible scenario is that the low-scoring group

might have scored some items correctly by chance, while the high-scoring group still missed the items.

To decide whether to delete or keep items, we referred to the “alpha if item deleted” and compared the recalculated alpha with the original alpha of 0.343. Although items 2 and 6 were evaluated as poor items with the d -value of 0.133 and 0.106 respectively, we decided to keep them in our test in that “the alpha if item deleted” rather decreased to 0.317 and 0.328 for item 2 and 6, respectively. These figures were slightly smaller than the original alpha of 0.343 and thus deleting these items would not help to increase the Cronbach alpha for the reading test. The same thing was true for the rest marginal items so we decided to keep item 1, 8, and 12.

When it comes to such questionable items as 4, 5, 9 and 11, more analyses are necessary to examine why the d -values turned out to be negative. In item 4, an inference question students had to infer a meaning of an expression in a context, it turned out that the lowest scorer got this question correct, while the highest missed the question. No consistent pattern was found among the middle group. The “alpha if the item deleted” went up to 0.516, which was much higher than the original alpha of 0.343. Since the item was considered to harm the test reliability with a negative discrimination index, we decided to eliminate item 4. Item 5 was a detail question that asked students to find information explicated in the text. Although searching for the explicit information was assumed to be an easy type of question, complex sentence structures of the text might have confused many students. The p -value of this item was 0.333, meaning that the question itself was too difficult so that only four students out of twelve scored

correctly. Since the “alpha if the item deleted” increased to 0.461, we decided to delete the item. Item 9 was another detail question that asked students to find the pronoun referent within a paragraph. Item 11 was a gist question that asked about the main idea of a paragraph. The negative *d*-value of these two items seemed to be due to their high *p*-values. In other words, the *p*-values of 0.8333 and 0.9167 for each item suggest that most of the students scored them correctly and thus the high and low groups were not properly distinguished. Given that the test was an achievement test and 70% of the students were expected to answer the questions correctly, we decided to keep the items despite the negative discrimination indices. Furthermore, the “alpha if the item deleted” for item 9 and 11 amounted only to 0.367 and 0.396, respectively. These figures were only a little larger than the original alpha of 0.343, compared to items 4 and 5 with the “alpha if item deleted” of 0.516 and 0.461. Based on these considerations, we decided to keep items 9 and 11.

All in all, we finally eliminated item 4 and 5 from the reading test and calculated the new Cronbach alpha (Table 10) and the new *SEM* (see Table 11). Consequently, the Cronbach alpha went up to 0.599 from 0.343. This implies the degree to which the items that relate to each other became higher, subsequently increasing the internal consistency. Likewise, the new *SEM* decreased to 1.324 from 1.551, which may also evidence the increased internal consistency. Taking participant #3 for example again, her raw score was 7, but the score was to vary between 4 and 11 (3.127 and 10.101 rounded due to the dichotomous scoring) when the *SEM* was 1.551 with a 95% confidence interval (± 2 *SEMs*). Now, with the recalculated

SEM of 1.324, her score would fall in between 5 and 10 (4.353 and 9.647 rounded due to the dichotomous scoring), if the test were repeated. Since the expected range between the lowest and the highest score with a 95% confidence interval (± 2 *SEMs*) slightly decreased from 7 (11-4=7) to 5 (10-5=5) with the new *SEM*, it seems safe to say that the internal consistency of this test improved, though the range of 5 could be still large for this short test with the total score of 12.

Table 10
Internal Consistency Coefficient Revised

Cronbach Alpha	Number of Items
0.599	10

Table 11
Standard Error of Measurement for the Reading Test Revised

$$*SEM = S \sqrt{1 - r_{xx}}$$

$$SEM = 2.09 * 0.6332 = 1.324$$

*where S = standard deviation (retrieved from the descriptive statistics) and *r_{xx}* = reliability estimate for the test, which is equal to the Cronbach's alpha coefficient.

Distractor Analysis

We also performed a distractor analysis to evaluate the quality of the individual items and to see whether they correctly discriminated the high group from the low group. The discrimination index was calculated by comparing the high 27% group and the low 27% group in their responses to the key answers and other distractors. The three top-scoring students were separated from the three bottom scoring students, as three is approximately 27% of twelve. Those who scored 9 and 10 points were selected as the high group, while those who scored 3, 5 and 6 were treated as the low group. Since there were four students who received 6, one was

randomly selected among the four and consistently used for the distractor analysis across different items. To calculate the discrimination index, the number of the high-scoring students that answered the item correctly was subtracted by the number of the low-scoring students that answered the same item correctly and then divided by the number of the high group students.

The value of this index is scaled from -1 to 1 ; the value of 0 indicates that there is no discrimination. The ideal value for the key answer is 1 or positive at least, while the value for distractors should be -1 or negative. The formula to calculate the discrimination index is presented in Table 12.

Table 13
Distractor Analysis for Item 4

Question Type	Answer	High 27% $N=3$	Low 27% $N=3$	Total Count	Total %	Discrimination Index	Difficulty Factor
Key	a	0	0	0	0	0	
Distractor	b	4	0	4	33	.33	0.667
Distractor	c	2	3	8	67	-.33	
Distractor	d	0	0	0	0	0	

Item 4 was a negatively discriminating item, and thus we decided to delete it. The key answer was *c* and around 67% of the students answered this question correctly. The discrimination index shows that no one chose the distractor *a* and *d*, meaning that these distractors did not function well as intended. The entire low 27% group got this question correct, while one student from the high 27% group chose the distractor *b*. Consequently, the key answer turned out to be negatively discriminating, while the distractor *b*

Table 12
The Formula for The Discrimination Index for The Distractor Analysis

D represents the discrimination index:
 Nch stands for the number of the high-scoring students who got an item correct, Ncl means the number of the low-scoring students who got the item correct and Nh means the number of the high group students, the formula for the discrimination index is, $D = (Nch - Ncl)/Nh$

Item 4 (Inference) with a d -value of -0.417 and item 10 (Detail) with a d -value of 0.706 were chosen for the distractor analysis to investigate what led to the discrepancy. Based on the formula above, the distractor analysis for item 4 was summarized in Table 13.

positively discriminating. This was a rather undesirable outcome in that ideally the discrimination index for the key answer should be a positive value or even 1 at the highest, while that of distractors should be a negative value or even -1 at the lowest. The undesirable function of the key answer and distractors in item 4 might have contributed to the negative d -value of -0.417 and the decreased internal consistency of the test. To improve the quality of the test, distractor *a* and *d* need to be replaced with more attractive

distractors; further revision is necessary to make the key answer *c* positively discriminating and make the distractor *b* negatively discriminating. Otherwise, it seems preferable to delete item 4 to

increase the test reliability. Now we turn to item 10, which had a *d*-value of 0.706 and a *p*-value of 0.250. Table 14 summarizes the results.

Table 14
Distractor Analysis for Item 10

Question Type	Answer	High 27% N=3	Low 27% N=3	Total Count	Total %	Discrimination Index	Difficulty Factor
Key	A	3	0	3	25	1.00	
Distractor	b	0	0	1	8.3	0.00	
Distractor	c	0	2	5	42	-0.67	0.2500
Distractor	d	0	1	2	17	-0.33	
Other				1	8.3		

Although this item was evaluated as an extremely difficult item with the *p*-value of 0.250, the distractor analysis revealed that it properly discriminated the high group from the low group, with well-devised distractors. The entire high-scoring group chose the key answer, while the low and the medium group selected other distractors. Consequently, the discrimination index of the key answer turned out to be 1, meaning that the item perfectly distinguished the high group from the low group. Distractor *c* was the most attractive, in that 42% of the students responded to it, and 40% of the respondents were from the low group. Distractor *b* and *d* also appealed to around 8% and 17% of the students respectively, but not to any in the high group. That being said, all distractors seem to have reasonably served their purpose.

Evidence for Construct Validity with the MC Task

Finally, the correlations among reading variables were examined to assess the construct validity of the MC items. Construct validity pertains to the question of the extent to which a test measures the underlying psychological

constructs of the test. Earlier in the paper, we decided to have gist, detail and inference variables to estimate the reading construct. That is, the three variables should be correlated with one another, as they all measure the same underlying construct. The Pearson product-moment correlation was computed; the range of the Pearson correlation coefficient ranges from +1 to -1. A positive value indicates a direct, linear relationship between the variables while a negative value indicates an inverse relationship. According to Brown (2005), there is a high correlation between the two variables when the coefficient equals to 0.75 or above, a moderate correlation when it falls between 0.5 and 0.74, a low correlation when it comes between 0.25 and 0.49. If the coefficient is below 0.25, it is safe to say that the variables are uncorrelated. When the correlation coefficient is close to 0, in either a positive or a negative figure, it indicates little or no correlation between the variables. Using these standards, we summarized the correlation analyses in Table 15.

Table 15
*Correlation Matrix between Variables
for the Reading Test (K=12, N=12)*

Scale	Gist	Detail	Inference
Gist	1		
Detail	0.142	1	
Inference	0.369	0.367	1

Note: * indicates significance at the $\alpha=0.05$ level (2-tailed)

A low correlation of 0.367 was found between inference and detail; another low correlation of 0.369 between inference and gist; near-zero correlation of 0.142 between detail and gist. Such a low or no correlation among reading variables refutes the sound construct validity of the test. In light of the generalizability of the correlation coefficient, each correlation coefficient turned out to be statistically insignificant. In considering that the observed correlations were smaller than the critical value of 0.576 ($df=10$) at the 0.05 level, we cannot rule out the possibility that the results presented in Table 15 were possibly due to chance.

Since items 4 and 5 were judged to have depressed the test reliability with the lowest d-value, we took out the two items from the analyses. After the deletion of the two items, the correlations among variables of the revised test were recalculated as shown in Table 16.

Table 16
*Correlation Matrix Between Variables
for the Revised Test (K=10, N=12)*

Scale	Gist	Detail	Inference
Gist	1		
Detail	0.310	1	
Inference	0.442	0.497	1

Note: * indicates that correlation is significant at the $\alpha=0.05$ level (2-tailed).

As a consequence, the magnitude of correlations among the variables slightly increased. A somewhat moderate correlation was found between inference and detail with the coefficient of 0.497. Still, a low correlation was estimated between gist and inference with that of 0.442 and between gist and detail with that of 0.310. Again, the correlational evidence among reading variables from the revised test was not sufficient to verify the construct validity of the reading test. We cannot guarantee the generalizability of the test result, since the observed correlations were not statistically significant at the 0.05 level.

Discussion and Conclusions

The purpose of this study was to demonstrate how to design the reading test, analyze the results, and evaluate the quality of the test. Given that it was an achievement test, its purpose was to measure the extent of learning or mastery within a specific instructional domain. Based on the theoretical model of the reading construct and the course syllabus specific to the A2 evening class, we developed 12 MC items for the reading test. We expected our test to correctly measure the underlying construct of reading ability. Although the topic of the passage was intended to correspond to the class theme, we tried to make the items not susceptible to their topical knowledge; to answer

questions, test-takers needed to closely read the passage.

Overall, the reading test turned out to be somewhat difficult for the students, in that the means, medians and modes of the construct did not meet the general standards of those of an achievement test. Even so, the test scores were rather normally distributed, indicating that the participant group was not as homogenous as we expected in terms of their reading ability. The results might suggest that the CEP placement test failed to place them according to their true language abilities, thereby calling for test improvement. Otherwise, it may also be that the participants were not motivated enough to do their best on the examination. All of the participants were adult ESL learners with a high level of general education, having at least a bachelor's degree. They voluntarily attended the CEP program to develop their general English ability, and hence they might not have felt much pressure about taking the test.

When it comes to evaluating the reliability and the construct validity of our reading test, by performing the item analysis, nine out of twelve items were evaluated as either marginal, poor, or even negatively discriminating items in our pilot test. Taking into account the "alpha if item deleted" and the p-values, we decided to delete two items. As a result, the Cronbach's alpha for the reading test increased, but still no statistically sufficient evidence was found for the construct validity. Thus, the reading test might not have been as successful in correctly measuring the underlying reading construct.

The undesirable outcome of the reading test seemed partially due to its elicitation method. While a writing or speaking task is a relatively direct test task, where test-takers are required to

do the actual skill, the MC items are devised to indirectly assess the intangible construct, reading ability. Thus, it is questionable if such items can actually tap into test-takers' true reading ability. Murphy et al. (1998) also indicates the fragility of the evidence surrounding reading assessment. Given that reading itself is a "complex and multifaceted process (p. 6)," it must be extremely challenging to access the abstract construct precisely.

Limitations

One of the main limitations, as mentioned several times earlier, was the small number of participants in the study (N=12) and limited number of items given on the test (K=10). These limited numbers could have been a factor that restricted evidence for the validity and generalizability of the test. Moreover, the range of ability among the participants was presumably rather narrow in that they were in the same level of CEP classes, limiting the variability of possible scores. A small range of variability can depress the correlation coefficients, and as a result, bring down test validity and generalizability.

Another limitation of the test is that, although the test items were created based on the CEP course syllabus to measure the participants' level of achievement, their scores did not reach a level that is generally expected in an achievement test. More specifically, an achievement test generally brings about an average score of 70% (which is also the cut-off score for CEP students when they advance to the next level), whereas our reading test average was only 56.25%. This figure could mean that the participants performed poorly overall, but on the other hand, it can also imply that the difficulty level of the overall test was rather high for the participants, or

even that the test was not an adequate representation of what they learned up to the mid-term exam.

There are some possible improvements we would make to the process of this project were we to administer it again. First of all, we would try to adjust the difficulty level of test, double-checking whether the items accurately reflect the course contents so that it would better serve as an achievement test. Closer communication with the teacher during the process of the test creation could help in carrying out this goal. It would also be helpful to administer a trial test with the items or have peers review the items to receive specific feedback before using the test. Finally, having a larger pool of participants would definitely help to obtain more reliable statistics when analyzing the test results.

References

- Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. Charles Alderson & A. H. Urquhart (Eds.), *Reading in a Foreign Language* (pp. 1–27). London: Longman.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Anderson, N. J. (1999). *Exploring second language reading: Issues and strategies*. Boston, MA: Heinle & Heinle Publishers.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460–472.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25, 133–150.
doi:10.1017/S0267190505000073
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. New York, NY: McGraw-Hill.
- Carrell, P. L., & Eisterhold, J. C. (1983). Schema Theory and ESL Reading Pedagogy. *TESOL Quarterly*, 17(4), 553–573.
- Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19(4), 727–752.
- Carver, R. (1997). Reading for one second, one minute, or one year from the perspective of reading theory. *Scientific Studies on Reading*, 1, 3–43.
- Coady, J. M. (1979). A psycholinguistic model of the ESL reader. In R. McKay et al. (Eds.), *Reading in a second language* (pp. 5–12). Rowley, MA: Newbury House.
- Cohen, A., & Upton, T. (2007). ‘I want to go back to the text’: Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209–250.
- Cummins, J. (1991). Conversational and academic language proficiency in bilingual contexts. *AILA Review*, 8, 75–89.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Ebrahimi, S. S. (2012). Reading strategies of Iranian postgraduate English students living at ESL context in the first and second language. *International Conference on Education and Management Innovation*, 30, 195–199.
- Educational Testing Services. (2000). Test of English as a Foreign Language. Retrieved from

- <http://www.ets.org/Media/Research/pdf/RM-00-04.pdf>
- Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research*, 65(2), 145–190.
doi:10.3102/00346543065002145
- Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahwah, NJ: Erlbaum.
- Goodman, K. S. (1973). *Theoretically based studies of pattern of miscues in oral reading performance (Final Report Project No. 9-0375)*. Washington, DC: US Department of Health, Education and Welfare, Office of Education, Bureau of Research.
- Goodman, K. S. (1994). Reading, writing, and written texts: A transactional socio-psycholinguistics view. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1093–1130). Newark, DE: International Reading Association.
- Grabe, W. (1991). Current development in second language reading research. *TESOL Quarterly*, 25(3), 375–406.
- Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. Harlow, England: Pearson Education.
- Grabe, W. (2004). Research on teaching reading. *Annual Review of Applied Linguistics*, 24, 44–69.
- Hoover, W. A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.
- Judith, R. (1995). Student responses to reading strategies instruction. *Foreign Language Annals*, 28(2), 262–273.
- Khalifa, H., & Weir, C. (2009). Examining Reading: Research and practice in assessing second language reading. *Studies in Language Testing (Vol. 29)*. Cambridge, England: UCLES/Cambridge University Press.
- Kitao, S. K. (1989). *Reading, schema theory and second language learners*. Tokyo: Eichosha Shinsha.
- McGinley, M. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27(3), 226–248.
- Murphy, S., Shannon, P., Johnston, P., & Hansen, J. (1998). *Fragile evidence: A critique of reading assessment*. Mahwah, NJ: Erlbaum.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8, 293–316.
- Paris, S., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Literacy Research*, 13(1), 5–22.
doi:10.1080/10862968109547390
- Savery, N. (2012). Targeted reading comprehension strategies instruction for raising reading levels in tertiary contexts. *Journal of Academic Language & Learning*, 6(1), 32–47.
- Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, 29(4), 431–449. doi:10.1016/S0346-251X(01)00039-2
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford Press.

Appendix A
Test Task Specifications

	Task 1 Multiple choice
SETTING	
Physical characteristics	Location: room HM 136 at Teachers College. Noise level: low to moderate depending if the door is open. Temperature and humidity: cool and moderate in humidity. Seating conditions: each test taker has his/her own seat in an auditorium type classroom. Lighting: well lit. Materials and equipment an degree of familiarity: pens or pencils, paper provided, students refer to clock on front wall to keep time.
Participants	The CEP teacher and the students
Time of task	During class hours at 7 PM on Thursday, October 23, 2008.
INPUT	
Format	
<i>Channel</i>	Visual
<i>Form</i>	Language
<i>Language</i>	Target: English as a second language
<i>Length</i>	Instructions: one to two sentences, Reading passage: ten paragraphs
<i>Type</i>	Item: elicit selected response
<i>Speededness</i>	Unspeeded
<i>Vehicle</i>	Live
Language characteristics	
<i>Organizational characteristics</i>	
Grammatical	Vocabulary: general Morphology and syntax: standard English Graphology: typewritten
Textual	Cohesion: cohesive Organization: focused discussion and analysis
<i>Pragmatic characteristics</i>	
Functional	Ideational, manipulative, and heuristic
Sociolinguistic	Dialect/variety: standard Register: formal Naturalness: natural Cultural references and figurative language: related to topic
Topical characteristics	Restricted: education issues in the US presidential candidate debate

EXPECTED RESPONSE	
Format	
<i>Channel</i>	Visual
<i>Form</i>	Non-language; circling the correct letter
<i>Language</i>	Target: English as a second language
<i>Length</i>	Short: 12 MC items
<i>Type</i>	Selected response
<i>Speededness</i>	Generally unspeeded
Language characteristics	
<i>Organizational characteristics</i>	
Grammatical	Vocabulary: general Morphology and syntax: standard English Graphology: circled responses
Textual	Cohesion: cohesive Organization: extended discussion and analysis
<i>Pragmatic characteristics</i>	
Functional	Ideational and heuristic
Sociolinguistic	Dialect/variety: standard Register: formal. Naturalness: natural Cultural references and figurative language: related to topic
Topical characteristics	Restricted: education issues in US presidential candidate debate
RELATIONSHIP BETWEEN INPUT AND RESPONSE	
<i>Reactivity</i>	Non-reciprocal
<i>Scope of relationship</i>	Broad to work with the general gist and inference questions. Narrow to work with the vocabulary and grammar in context questions
<i>Directness of relationship</i>	Direct

Appendix B

Test-taker Survey**A. Personal characteristics**

1. Age:
2. Gender: M / F
3. Nationality:
4. Native language:
5. How long have you been in the United States? _____
6. What is the level of education that you completed at the most recent years?
 - a. elementary
 - b. secondary
 - c. undergraduate
 - d. graduate
 - e. post-graduate
7. Are you planning to go to college or graduate school, or find a job in the United States? Y / N

B. Topical knowledge

1. How often do you read an American newspaper?
 - a. everyday
 - b. every other day
 - c. once a week
 - d. once a month
 - e. never
2. How many hours do you spend in reading a newspaper?
 - a. less than half an hour
 - b. half an hour
 - c. one hour
 - d. two hours
 - e. more than two hours
3. Are you interested in 2008 US presidential election?
 - a. very much
 - b. interested
 - c. only a little interested
 - d. not interested
4. What did you major in, if you have a bachelor degree? _____

C. Levels and profiles of language knowledge

1. How much time have you spent studying English (in a secondary or post secondary school)? _____
2. Have you ever taken any standardized English exam (e.g., TOEFL, TOEIC) before? If so, which test was it and what was your score? _____

D. Possible affective responses to taking the test

(5 = strongly agree , 1 = strongly disagree)

I was nervous while taking the mid-term exam	5	4	3	2	1
I am familiar with the types of questions in the reading section.	5	4	3	2	1
I am familiar with the type of writing question	5	4	3	2	1
I felt the level of the reading questions was difficult	5	4	3	2	1
I felt the level of the writing question was difficult	5	4	3	2	1

E. Reflecting on the test questions.

1. How did you solve the following question in the reading section?
 - a. I already knew this information from the media (e.g., newspaper or TV).
 - b. I skimmed the reading passage to find the information.
 - c. I just guessed randomly.
 - d. Other ways _____

9. In his presidential campaign on the issue of education, John McCain:

- a. disagrees with the idea of NCLB.
- b. suggests more grants for preschool programs.
- c. wants to reward high-achieving teachers with federal money.
- d. plans to increase government funding for independent schools.

2. How did you solve the following questions in the reading section?
 - a. I already knew the meaning of the word before taking this test.
 - b. I inferred the meaning from the content of the passage.
 - c. There is a similar word in my first language.
 - d. I just guessed randomly.
 - e. Other ways _____.

6. In line 12, what does "glum" mean?

- a. puzzling
- b. convincing
- c. discouraging
- d. self-explaining

12. What is the meaning of "diluted" in line 46?

- a. less effective
- b. risky to carry out
- c. more troublesome
- d. difficult to clean up

3. How do you rate your reading ability in your first language?
 - a. Advanced
 - b. High-intermediate
 - c. Low –intermediate
 - d. Beginner

4. How do you rate your writing ability in your first language?
 - a. Advanced
 - b. High-intermediate
 - c. Low –intermediate
 - d. Beginner

5. In the writing section, was it helpful to have the planning chart before writing the essay?
 - a. Very helpful
 - b. Somewhat helpful
 - c. Only a little helpful
 - d. Not helpful

Appendix C

Mid-term Evaluation for CEP A2 Evening Class

Name: _____

Instructor: Abbi Leman (A2 Evening)

Date: Oct. 23, 2008

READING SECTION

You have 30 minutes to complete the following reading tasks.

Directions: Read the passage. Circle the correct letter.

“OUR nation is at risk. Our once unchallenged pre-eminence in commerce, industry, science and technological innovation is being overtaken by competitors throughout the world.” So reported an education commission in 1983. That report was a turning point for American schools, helping spur a wave of reform. But 25 years later the state of American education is in a muddle.

5 In some ways its public schools have improved. America’s nine-year-olds scored 22 points higher on a national maths test in 2004 than they had in 1982. But in many areas America still languishes, as described in a recent report by Ed in ’08, an advocacy group. The percentage of 17-year-olds with basic reading skills has dropped, from 80% in 1992, when the current test was introduced, to 73% in 2005. On the international stage, American students are doodling while others scribble ahead. The Organization for Economic Co-operation and Development has a glum statistic: in the most recent ranking of 15-year-olds’ skill in maths, America ranked 25th out of 30. Though America’s universities remain pre-eminent in the world, they have grown increasingly unaffordable. Barack Obama notes that between 2001 and 2010, two million qualified students will not go to university because they cannot afford it.

10

Efforts to move America forward have proceeded inconsistently. A federal bill, No Child Left Behind (NCLB) was passed with broad support in 2002, the culmination of a long push to set high standards and hold schools accountable for meeting them. It requires states to test students on maths and reading; science is being added. Schools that do not progress towards meeting state standards face financial sanctions.

15 But the law is hotly debated. George Miller, a Democratic congressman, calls NCLB “the most negative brand in America”—and he was one of the law’s architects. Teachers’ unions utter no four-letter word with more anger than NCLB. They say the law forces “teaching to the test”, that the sanctions are too strong and the carrots too small. Even those who still support the law find problems with it. NCLB, for example, does not chart a student’s progress.

20

Some states have set their standards very low. Some 90% of Mississippi’s fourth-graders were labeled “proficient” or better on a state reading test in 2007; only 22% were so described after a national test.

Unsurprisingly, advocates from all corners are trying to make education a main campaign issue. Ed

in '08 points out that many of the proposals from "A Nation at Risk" have been ignored: standards remain weak, few districts pay teachers by results and calls for a longer school year have gone disregarded. But despite a budget of \$60 million, Ed in '08's campaign has had little impact.

25 Mr. Obama is at least taking the problem seriously. His plans run the gamut, from grants for preschool programs to a \$4,000 tax credit for university fees. He is vague about NCLB, but has resisted calls to throw out the law. He suggests improving it through more sophisticated tests, measuring students' progress over time and giving schools more resources. In September he announced new plans to double federal funding for independent or "charter" schools. A separate "innovative schools fund" would help districts to create a portfolio of successful school types, including charters.

30 Perhaps most interesting are his plans for teachers. He would give extra money to districts that work with their unions to form "career ladders". These could include pay increases for a list of achievements, from teaching in hard-to-staff schools to lifting students' performance.

But a good scheme on paper may be diluted in practice. Negotiations over pay are messy at best.

For his part, Mr. McCain offers promising opinions but few details. He supports NCLB but has said little about how to strengthen its main tenets. He supports charter schools (like Mr. Obama) and voucher programs (unlike Mr. Obama, who is dead-set against them), but has said little about how he might expand them. His boldest ideas center around using federal money to let parents choose tutors and principals reward good teachers.

35

In the debate over how a president might help America's schools, a main obstacle is that, traditionally, it has not been his job to help them much at all. The national government provides less than 10% of total spending on schools. Indeed, states and cities continue to be the boldest innovators. Chicago is opening dozens of new schools, including charter schools, in its poorest areas. Cities such as Denver and New York now have schemes to reward teachers for their skill. The results there are mildly encouraging.

40 The two candidates offer different plans for how they might push these reforms along. Both, however, have largely overlooked the most obvious role. At the very least, the next president could help to create a better benchmark for student achievement. As Mississippi proves all too well, a state standard can be an elastic ruler.

What is the best title for the passage? - GIST

- a. Under NCLB, even strong schools falter
- b. Can the candidates fix America's decidedly mediocre schools?
- c. Can school equity be achieved with a larger education budget?
- d. Discrepancies between McCain and Obama over education policies

1. What is the author's overall tone in the passage? - INFERENCE

- a. Ironic.
- b. Neutral.
- c. Critical.
- d. Hopeful.

2. Why does the author mention the education commission in 1983 in the beginning?

-INFERENCE

- a. To point out the effects of American education on other social areas since 1983
- b. To emphasize that American education has been a problem for the past 25 years
- c. To give an example of the efforts that a government made to improve education
- d. To relate the event to the education policies that two presidential candidates suggest

3. What does it mean to "be in a muddle" in line 5? - INFERENCE

- a. be in mental stress
- b. lack attention to details
- c. be in a disorderly condition
- d. have no sense of responsibility

4. What is true according to the 2nd paragraph (lines 6–16)? - DETAIL

- a. Public schools in America have made overall improvement.
- b. America still has a relatively good international ranking in math skills.
- c. Math abilities of nine-year-old children enhanced significantly by 2004.
- d. Even American universities are falling behind in terms of academic competence.

5. Why does the author mention George Miller in line 22? - INFERENCE

- a. To reveal the controversy of the NCLB debate
- b. To give a specific example of one limitation of NCLB
- c. To provide evidence of how strongly NCLB is opposed
- d. To suggest that Miller would be able to improve NCLB

6. Which of the following would best replace the word “carrots” in line 25? - INFERENCE
- a. support
 - b. rewards
 - c. challenge
 - d. standards
7. In his presidential campaign on the issue of education, John McCain: - DETAIL
- a. disagrees with the idea of NCLB.
 - b. suggests more grants for preschool programs.
 - c. wants to reward high-achieving teachers with federal money.
 - d. plans to increase government funding for independent schools.
8. In line 39, what does “it” refer to? - DETAIL
- a. NCLB
 - b. the law
 - c. the gamut
 - d. the problem
9. Based on the facts in the 5th paragraph (lines 28-30), which of the following correctly rephrases “a state standard can be an elastic ruler” in line 62? - DETAIL
- a. Standards set within the state can be misleading
 - b. The state legislators can be flexible in law making
 - c. States can set standards that increase student performance
 - d. Sometimes states can measure students upon a rigorous standard
10. The main point of the 10th paragraph (lines 53-58) is: - GIST
- a. Future American president needs to allow more budgets for schools.
 - b. Efforts for better education have been made mostly at the state level.
 - c. The national government is planning to give teachers more incentives.
 - d. Many states are against education policies that the federal government suggests.
11. What is the best conclusion of the passage? - GIST
- a. Teachers and government officials must all cooperate toward improving the effects of NCLB.
 - b. Enhancing the quality of education in America will be a major job for the next president.
 - c. States and cities should take the more initiative role to improve schools and student performance.
 - d. The first step towards reform can be made if the national government increases the funds for education.