The Korea TESOL Journal
Volume 1, Number 1   October 1998

Edited by **Thomas Farrell,** National Institute of Education, Singapore
Assistant Editor **Carl Dusthimer,** Hannam University, Taejon, South Korea

Contents

Editor's Note

Welcome to the first issue of *The Korea TESOL Journal*. I am delighted to edit the first issue of *The Korea TESOL Journal*. It may seem strange to readers teaching in Korea that the editor is not living and teaching in Korea. Well, I was there living and teaching in the country for 16 years, up to June 1997. During that time I was in the background of Korea TESOL working mostly with the current president, Carl Dusthimer. During those years we had discussed the need for an academic journal that would make a contribution to the development of English teachers in Korea. This journal would include research that was carried out in Korea and the Asian region, but would be more than just a place to publish papers. It would also be a window into the rich and diverse research that was and is on going in Korea and the region. We felt that this research especially by practicing EFL teachers in Korea has not been adequately represented in other newsletters or past Korea TESOL journals. This research will make a contribution to the TESOL field for, as I still believe, Korea has a lot to offer in terms of research into language teaching and learning. For me this dream has come true with this first issue of the journal. What better time for the inaugural issue than to appear with the Korea TESOL Conference in Teacher Development—the focus of the journal. I hope you enjoy this first issue and contribute to further issues.

For this issue articles were solicited from various people around the world, especially teachers who had a connection with Korea in the past and carried out research in Korea or trained teachers to teach in Korea.

- The first article is an example of the latter. Jerry Gebhard has trained and developed many teachers in the United States at Indiana University of Pennsylvania (IUP), myself included, and has a keen insight into teacher development as it concerns teachers of English in Korea. This feature article defines teacher development, considers areas of teaching that can be developed, describes ways teachers work on the development of their teaching, and raises questions about teacher development as it relates to Korea.

- In the second article Matt Peacock describes an investigation into the correlation between learner answers on post-class materials evaluation questionnaires, filled out anonymously, and comments made by the same learners during subsequent interviews. The aim of the investigation was to determine the extent of learner truthfulness during the interviews, and thereby to make a preliminary assessment of the correctness of the suggestions of Oller. Results indicate that the mean agreement between the questionnaire response and interview response for individual learners was fairly high.

- The third article examines to what extent one teacher, in this case the author, Peter Nelson, reliably and systematically evaluated major characteristics of connected speech. Its focus is not free or structured conversation, but student readings of two passages which offered a variety of phonotactic features and facilitated differentiation among them. As presented, then, the students' pronunciation characteristics should not be considered as proxies for their performance in conversation, since intonation patterns, elision, linking and other features normally associated with free speech tend to be different when students are reading. In effect

the methodological setting is deliberately limited, focusing on a teacher's listening skills in an artificial context.

- In the fourth article, Graeme Cane looks at the ways in which conversation skills are presented in EFL textbooks and courses, and suggests some alternative sources and approaches in teaching these skills.

- Finally, in the fifth article, Tom Farrell maintains that EFL teachers can make their own language placement tests that best suit their needs and context. His original test, the oral placement test (the OPT) was analyzed by another researcher ten years on. The findings show that the test, which is still in use, is generally reliable and is still valid.

Thomas Farrell

Second Language Teacher Development and Korea

JERRY G. GEBHARD
Indiana University of Pennsylvania

Teacher educators have shown interest in second language teacher development for a number of years. For example, several second language teacher educators emphasized that teachers need to gain skills to make their own informed teaching decisions rather than blindly following the latest teaching methods or fashions. During the past decade, interest in teacher development has blossomed in such areas as exploring our beliefs and teaching practices through observation, reflecting on teaching, exploring the meaning of teacher development as opposed to prescriptive training, and exploring the meaning of cooperative development. The purpose of this article is to define teacher development, consider areas of teaching that can be developed, describe ways teachers work on the development of their teaching, and raise questions about teacher development as it relates to Korea.

Teacher educators have shown interest in second language teacher development for a number of years. For example, several second language teacher educators emphasized that teachers need to gain skills to make their own informed teaching decisions rather than blindly following the latest teaching methods or fashions (Clarke, 1982; Fanselow 1977; Fanselow and Light, 1977; Larsen-Freeman, 1983). During the past decade, interest in teacher development has blossomed. For example, Fanselow (1987, 1988, 1992) has introduced us to a process of exploring our beliefs and teaching practices through observation. Freeman (1989) has brought our attention to the meaning of teacher development as opposed to prescriptive training. Edge (1992) has shown us the meaning of cooperative development. Richards and Lockhart (1994) have introduced us to reflective teaching as a means to develop. Likewise, Richards and Nunan (1990) put together a now classic edited volume on second language teacher education. Many of the chapters in this book focus on teacher development and ways teacher educators can provide chances for language teachers to develop as teachers. Two other books, which include conference papers on teacher education and development, are Flowerdew, Brock, and Hsia (1992) and Li, Mahoney, and Richards (1994). More recently, a special issue of *TESOL Journal* (Vol. 7, No. 1, 1997) has been devoted to professional development and I have put together a book with Robert Oprandy (Gebhard & Oprandy, in press) in which we go into detail about ways teachers can gain develop as teachers through observation, action research, keeping a journal, talking, and reflecting on personal experience.

The purpose of this article is to define teacher development, consider areas of teaching that can be developed, describe ways teachers work on the development of their teaching, and raise questions about teacher development as it relates to Korea.

TEACHER DEVELOPMENT DEFINED

In order to define teacher development, allow me to contrast *teacher development* with *teacher training* (Freeman, 1989; Larsen-Freeman, 1983; Lange, 1990; Richards, 1987, 1989). When teachers are in college and undergo a *training* approach to their education as teachers, they take a series of courses, many with lectures. They read about and listen attentively to the professors talk about such

topics as language, linguistics, language acquisition, methodology, and materials design. They take tests and write papers. If their program has a training emphasis, then during their practicum or student-teaching experience, student-teachers are assigned a cooperating teacher or a supervisor who models teaching, prescribes the "best" ways to teach, and has them practice teaching using techniques repeatedly until they master them. For example, they might practice waiting longer after asking questions, or practice working through a set of prescribed procedures for teaching a reading lesson.

However, the problem is they only learn how to teach in the "prescribed" way, and they have no real way to gain awareness of their beliefs about teaching and learning, attitudes towards students, teaching, and related matters, nor ways to reflect on better understanding their teaching and the consequences it has on the students.

Teachers who experience a *development* focus also take courses. However, course work have fewer lectures and tests and more projects and portfolios that are designed to get them to think about teaching and themselves as teachers. For example, teachers might keep a dialog journal in which they keep a teaching journal (Bailey, 1990; Brock, Yu, & Wong, 1992, Gebhard & Oprandy, in press). They might participate in activities that allow them critically to reflect on their teaching, such as what Johnson (in press) calls "Tasks for Critical Reflection on Teachers' Knowledge." Or, they can be asked to write autobiographies to discover such things as how their past learning experiences affect who they are as teachers (Bailey et. al., 1996; Johnson, in press). Or, tasks might be used to teach teachers how to categorize and re-categorize their knowledge (Ellis, 1987; Fanselow, 1992), relate personal experience to teaching and being a teacher (Gebhard & Oprandy, in press), or question their attitudes (Pennington, 1990). They might also design materials and media they predict will be useful to their future teaching contexts, then put on a Materials/Media Fair for their peers, friends, and invited guests (Gebhard, 1993). If they are student-teaching, they might work on collaborative action research or a professional project (Wallace, 1996).

Training has been characterized as a "deficiency view" of educating teachers (Freeman, 1989; Lange, 1990; Richards, 1987). Teacher educators approach education as if the students have little previous knowledge and few beliefs or attitudes about teaching. In contrast, those who believe in a development approach see each teacher as someone with unique knowledge, beliefs, and attitudes. The teacher educator's job is to provide chances for teachers to gain awareness of what they already know, believe, and feel about teaching and themselves as teachers, as well as provide opportunities for them to gain new knowledge, and develop their beliefs and attitudes.

I want to emphasize that teachers don't have to go through a formal preservice teacher education program to learn to work on developing their teaching. Inservice workshops and teacher self-development groups can also offer teachers chances to experience different teacher development activities and to understand how they can work on developing their teaching on their own. And, some teachers have decided on their own, without the stimulus of a teacher education workshop, that they need to constantly work at developing their teaching. For example, Farrell (1996) describes a group of EFL teachers in Korea who decided to develop their teaching through observation and reflection. Such teachers agree with Lange (1990), who sees development as "a process of continual intellectual, experimental, and attitudinal growth" (p. 250). Development is a process of evolving as a teacher, of the continual unfolding of beliefs and teaching practices throughout a teacher's career.

WHAT TEACHERS CAN DEVELOP

As Freeman (1989) points out, teachers can develop knowledge, skills, attitudes, and beliefs. Knowledge, for the teacher, includes "what is being taught (the subject matter); to whom it is being taught (the students - their backgrounds, learning styles, language låvels, and so oné; and where it és being taught èthe sociocultural, institutional, and situational contexts)" (p. 31). Skills, on the other

hand, are what the tåacher has to beàable to do in the classroom. Foò example, teachers have to do such things as present materials,àexplain conceptó, group learneró, give instructions, and keep students on task. As Freeman makeó clear, "taken together, these ãonstituents - kîowledge and kno÷-how, or skillsímake up what is often referred to as the knowledge base of teacèing" (p. 31).

Kîowledge and skills are importanô, but as Freeman also explains, it is also possible for teachers to develop attitudes and beliefs about teaching. He defines attitudes as "the stance one adoptó toward oneselfì the activity of teaching, and ôhe learners one engages in the ôeaching/learning process" (p. 32). Here are a few examples: "These students are so lazy that I really dread coming to class." "My job is to coach students to pass the college entrance exam." and "I'm not really a teacher. This is something I do to pay the bills".

Beliefs, like attitudes, are important because they are often the underlying force behind what and how we teach. Richards and Lockhart (1994) offer five categories of beliefs, including beliefs about (a) English, (b) learning, (c) teaching, (d) program and curriculum concerns, and (e) language teaching as a profession. All of these areas are open for development, as can be reflected in the questions teachers ask. For example, while developing beliefs about teaching English, some teachers ask: "Is it important for students to pronounce the language like a native speaker?" When addressing beliefs about learning, I have heard teachers ask: "Can students learn to be communicatively competent in a foreign language classroom?" Likewise, when addressing beliefs about teaching, teachers have asked: "How important is it to teach grammar?" "How can I get learners to take on more responsibility for their own learning?" "What are the principles that guide my way of relating to students?" And, questions which reflect their beliefs about program and curriculum concerns include: "What value does the text have for me? For the students?" "Should the curriculum be geared to passing a test?" Finally, I have heard teachers ask questions which reflect their beliefs about the language teaching as a profession: "Do I see myself as a true professional? Do the administrators who hire me?" and "How can I develop as a professional language teacher?"

HOW TEACHERS CAN WORK ON DEVELOPING THEIR TEACHING

Earlier in this article, I introduced a variety of activities that teachers use to develop their knowledge, skills, beliefs, and teaching practices. Here I discuss three kinds of activities that teachers can use to develop. These include development through self-observation, observing other teachers, and doing action research.

Development through Self-Observation

One way to develop as teachers is through self-observation. As I also write about elsewhere (Gebhard, 1996), teachers can gain a great amount of awareness about their teaching by systematically observing what goes on in their own classrooms. I suggest that teachers explore their own teaching through a cyclic process of reflecting and then acting on knowledge gained through reflection. The steps in the process include: (a) collecting samples of teaching, (b) analyzing the samples, (c) appraising or reflecting on the teaching represented by these samples, and (d) considering how the same lesson could be taught differently and designing the next lesson based on new awareness.

Collecting samples of teaching

There are a variety of ways to collect descriptive samples of what goes on the classroom. Video taping classroom interaction is one possibility. I prefer to audio tape the classroom interaction (after asking students for permission). The reason to tape the class is to have a way to analyze the teaching. A tape can be listened to (or viewed, in the case of video taping) over and over again. If

audio taping is done, I recommend teachers use small hand-held tape recorders. It can be turned on, set down, moved from place to place, even held in one hand while teaching, letting the tape recorder simply be an extension of the arm and hand.

This process of collecting samples of classroom interaction is more focused if there is a specific interest. Teachers who focus the observation tend to have particular kinds of questions in mind. Here are some examples from Gebhard (1996):

- What kinds of questions do I ask most often? Yes-no? Either-or? Wh-? Tag?
- What are the content of my questions? About the study of language? People's lives in general? Students' personal lives? Procedures? Other?
- How long do I wait after asking a question to get a response?
- How do I give instructions? How much time does it take? Do students know what to do after being given the instructions?
- How often do students speak their native language in class? When? What do I do when they use it?
- How do I praise students? What words do I use? What kinds of student behavior do I praise? How often do I praise? What don't I praise?

I want to emphasize that samples can also be collected randomly, with no specific focus and with the idea of simply exploring what goes on in the class. The value of simply collecting random samples of classroom interaction is the chance to get beyond preconceived notions about what is worth observing in the classroom. Sometimes, simply turning on a tape recorder, moving it around the classroom from place to place without giving it much thought, can result in some interesting discoveries about what actually goes on. For example, I put a tape recorder with a group of students who were working on a problem-solving task. Both the students in the group and I apparently forgot about the tape recorder. Later, when listening to the tape, I discovered a few things. First, students used more of their native language (Thai, in this case) than I had thought. They also went off task a number of times, talking about a variety of subjects (often in Thai).

Analyzing Samples of Teaching

There are different ways to analyze descriptions of classroom interaction, depending on the focus of the observation. Teachers can take notes while listening to the tape, make short transcripts, or tally behaviors. For example, if the interest is in students' reactions to instructions to do a small group task, then the teacher might listen to the recording while writing down the instructions. The teacher might also jot down notes on what the students did just after the instructions were given. Did students start the task right away? Ask each other what they are supposed to do? Ask the teacher to repeat the instructions? On the other hand, if the purpose is to gain awareness of the way the teacher treats students' spoken errors, the teacher might tally the number of times errors are corrected, categorize the kinds of errors that are treated, and make short transcripts which illustrate the way that the teacher and students interact when errors are corrected.

The time needed for an analysis depends on what the teacher wants to get out of the observation. In fact, it is not always necessary to spend a lot of time analyzing what goes on in the class. For example, let's return to the interest in how much native language the students use in the classroom and when it is used. To learn about this, the teacher could listen to the tape recording in the car or train while commuting. Simply listening can raise a lot of awareness. Or, if a little more interested, while listening to the tape, the teacher could jot down notes and tally how many times the students used their native language. If the teacher is deeply interested, he or she might work through the tape, studying conversations by making short transcripts of the interaction before students start using their native language to try to determine what triggers the switch from English to their native language.

Appraising or Reflecting on the Observed Teaching

The next step in the process is to appraise or reflect on the teaching. When doing self-observations of my own teaching, I like to ask these questions: How are opportunities possibly provided through the interaction for students to learn English? How are opportunities possibly blocked? Why do I teach the way that I do? What are my beliefs about teaching and learning? Does the interaction in the teaching I just observed reflect my beliefs? It is through searching for answers to questions like these that our beliefs about and attitudes toward teaching can be challenged and developed further.

Generating Changes and Exploring Ways to Teach Differently

To me, an important part of collecting and analyzing classroom interaction is not only to gain awareness of what is going on in the classroom, but also to generate alternative teaching behaviors based on this awareness. As such, I like to ask and answer questions such as these: What do I want to keep doing? What do I want to change? How can I bring about new consequences?

I like to follow John Fanselow's (1987, 1992) idea that we can explore teaching by trying the opposite of what we normally do. If a teacher discovers that she always teaches from the front of the classroom, she can explore by seeing what happens if she teaches from the back. If the teacher asks all the questions, this teacher could try teaching without asking any questions. What would happen? Maybe the students would ask questions. Maybe there would be lots of silence. Or, the teacher might have a spontaneous teaching idea, one that students react to in a positive way.

I also like Fanselow's idea that small changes can have big consequences. For example, I know a teacher who discovered she praised students all the time. She even praised them when they got wrong answers to questions. So, she decided to only praise them when they did outstanding work. This small change brought about a wonderful consequence. She believed that students worked harder to gain her praise and this resulted in students using English more in class.

The Observation Process Illustrated

To illustrate how the process of describing, analyzing, appraising/reflecting, and generating changes can work, here's a self-observation I did while teaching an American literature course in Hungary. Since I wanted to understand my questioning behaviors, I designed a tally sheet. I audio taped my class, and using a tally sheet (below), I later kept track of the targets of my questions (e.g., to an individual student or the whole class) and the content of each question (e.g. about students' lives, about people and places in general, or about the content of the reading selection). The following tally sheet records what I discovered (first published in Gebhard, 1991):

Content of Questions	To Individual	To Whole Class
Questions about students lives	//	
Questions about Hungarian culture		////////////
Questions about language	/	////////
Questions about reading material content		//////

I discovered that I asked twenty-eight questions during a twenty-five-minute time period, most of my questions were addressed to the whole class, and twelve questions were about Hungarian culture in general, eight about language, and six about the content of the reading. Upon reflection, I wasn't surprised that I asked mostly whole-class questions. However, I was surprised that I averaged over a question per minute! This discovery was very useful. It gave me the chance to reflect on my questioning behavior. As a part of my reflection, I thought about how discussions go on outside classrooms, how all the participants not only answer questions but also ask them and react to each others' responses. I was also able to see that my questioning techniques dominated class discussion and prevented students from raising their own questions and reacting to responses. Based on this

reflection, I decided to systmatically modify my questioning behavior. In the next seminar, I consciously asked less questions and attempted to achieve more discussion based on a single question. This small change created a more interactive classroom where the students asked me and each other questions, as well as reacted more to the content of the reading and to each others' comments.

Development through Observation of Other Teachers

Another way teachers can work on their development is to observe other teachers. The logic of observing others can be especially useful for some teachers. By observing others, as Fanselow (1988) points out, each of us has the chance "to construct, reconstruct, and revise our own teaching" (p. 116). He makes it clear that we can gain self-knowledge and self-insight by observing what teachers and students do in classrooms. He also makes the point that when we generate our own alternative ways to teach based on our observations of what others do, we are constructing our own knowledge. In short, we are developing our teaching by seeing our own teaching in the teaching of others.

The observation process for observing others is similar to the self-observation process. One advantage, of course, is that we can take notes and draw sketches, jot down quick samples of interaction, and tally behaviors as we observe, although I still like to tape some classes, however, only when the teacher and students approve.

I have known teachers who have been pleasantly surprised by how much they have gained from observing other teachers' classes. The first thing that most teachers ask is why teaching is such a "closed door" profession. Teachers do not visit other teachers classrooms very often, and teachers who do often say they gain a lot from the observation experience.

An Example of Observation of Other Teachers

Here is a documented example (Akamine, 1991) of what one teacher gained from observing other teachers. The teacher, a junior high school English teacher from Okinawa, was studying in the United States. She gained a strong interest in how she might further develop her skills at teaching vocabulary, and she decided to observe other language teachers to see how they did this. She began with a simple checklist of the usual ways vocabulary is taught (for example, by giving translations, definitions, sample sentences with the new word in it, synonyms and antonyms). She then visited ESL vocabulary building and reading classes. As she observed, she checked off those items that she observed that were on her checklist. But, she went beyond this. She also took notes on those techniques that teachers used that were not on her check list. She then added each new item to her list, and the list grew longer and longer the more she observed. Her one page check list eventually turned into five pages, and her ideas about how vocabulary can be taught, as well as her beliefs about learning vocabulary, changed considerably.

Development through Action Research

Another approach to developing our teaching is action research. Action research originated through the work of Kurt Lewin (1948), a social psychologist who brought together experimental approaches to social science research and the idea of "social action" to address social issues.

The Action Research Process

Action research, as discussed by such educators as Crookes (1993), Gebhard and Oprandy (in press), Kemmis and McTaggart (1982), and Strickland (1988) includes posing problems based on what goes on in teachers' own classrooms and lives and systematically working through the problems by creating and initiating a plan of action, as well as reflecting on the degree to which the plan works. Often action research is a community effort. The posed problem is discussed with other teachers who

offer their support and experience to help clarify and solve the problem. In addition, as Crookes (1993) emphasizes, action research can go beyond the confines of the teacher's classroom. When teachers discover that they share similar problems (i.e., large numbers of students across the program cutting classes, complaining about the curriculum, and not doing homework), it becomes an institutional or school problem, one which needs collaboration among teachers, students, and administration.

Action research usually follows a cyclic process which follows a series of steps. Below is a list of steps I have used and taught teachers to use. (See Gebhard and Oprandy, in press, chapter three, for elaboration on this process and example action research projects).

1. Identify an issue, interest, or problem.
2. Collaborate with others. Seek knowledge to clarify the issue, interest, or problem.
3. Plan an action to learn about the issue or interest or solve the problem. Report on the plan; use feedback to adjust the plan.
4. Implement the action.
5. Observe the action. (Observation techniques discussed earlier are relevant).
6. Report on the action taken and observation of the consequences.
7. Reflect on the observations and the report discussion.
8. Revise the plan.

An Example Action Research Project

Here is an example of an action research project. The teacher was teaching first semester Japanese at a university in the United States (Gebhard & Ueda-Motonaga, 1992). The aim of the course was to teach students to use Japanese for communication purposes to accomplish everyday things, for example, to shop, talk about the weather, and handle basic functions of language, such as apologizing and greeting. The teacher was assigned a text, *Japanese: The Spoken Language* (Jorden & Noda, 1987). The teacher and students followed the text in a routine way, starting class with practice dialogues, followed by pattern practice drills, and situational drills.

The teacher met with me, her supervisor on a weekly basis. We saw these meetings as a chance to learn more about teaching, and a project emerged. During the first meetings we studied short transcripts made from audio tapes of the teacher's class. Through analysis of the transcripts, she recognized a communication pattern going on in her classes, one which indicated the members of this class were not genuinely engaged in communication. She asked the majority of the questions and students only asked question when she told them to (e.g., Mr. Miller, ask Ms. Brown a question using the pattern in the book). She also mostly asked questions to which she already knew the answer (known as display questions), and that there was very little reaction to responses to questions from the students, something that frequently happens in everyday communication outside classrooms. She concluded that most class time was spent studying about Japanese, rather than learning to use Japanese to communicate, the goal of the course.

Based on this awareness, the teacher planned an action. To move toward a more communicative type of classroom interaction, she decided to use personal questions. Since she knew that many of the students went to a near-by city during spring break, she also decided to ask them about their trip, hoping that such questions would inspire students to communicate in Japanese. She also decided to bring a map of the city so that the students could point out where they went. She audio taped the classroom interaction to later study whether or not she was successful at getting students to communicate in Japanese through her personal questions.

As planned, she began the class by asking students in Japanese where they went during the spring break. A number, as predicted, gave the name of the near-by city. She then asked where in the city they went, and she brought the map out so that they could show her. Then, something new happened. Students started to ask each other questions and to react to each others comments in Japanese. They asked each other where different places on the map were, if they went shopping, how

expensive things were, and whether or not they had fun. It was the first time that students communicated with each other in Japanese.

By doing this action research project, with my support, the teacher had a way to develop her teaching knowledge, skills, attitudes, and beliefs. She was able to identify, talk about, and formally pose a problem in her teaching, plan and implement an action, and report on the project informally to me and formally through our publication. As a result, the teacher had lots of opportunities for development. She reported, for example, that she developed a stronger belief that language should be taught as communication rather than as grammar and translation practice. She also gained skills to do this, and gained a new attitude toward the importance of consistently paying attention to her teaching through observation and action research.

Teacher Development and Korea

In this article I have defined teacher development as a process of continual intellectual, experimental, and attitudinal growth, and as a process of evolving as a teacher, of the continual unfolding of beliefs, attitudes and teaching practices throughout a teacher's career. I also briefly discussed areas of teaching that teachers can develop, those of knowledge, skills, beliefs and attitudes, as well as pointed out a variety of teacher development activities. I discussed three kinds of activities in more detail, including self-observation, observation of other teachers, and action research.

But, I have not offered prescriptions about what teachers in Korea should do to develop or even stated whether teachers should be given opportunities and encouragement to develop. Rather, I would like to end this article with a few of the benefits of teacher development for educators in Korea to consider, as well as raise questions for teachers and teacher educators in Korea to think about.

One benefit of teacher development is that it adds to our abilities to reflect critically on teaching. As Richards (1990) makes clear, the more critical we can become, the more likely we can move from a level where we are guided by impulse, intuition, or routine to one in which our actions are guided by critical thinking.

Another benefit of development is that it can keep us interested in teaching. I know that in my twenty-three years as a language teacher and teacher educator I have constantly been interested in teaching and student learning. I believe part of the reason is because I have constantly paid attention to developing my teaching knowledge, beliefs, attitudes, and practices. The changes that concur with my development have made teaching very enjoyable.

Another benefit is that working on development of beliefs, attitudes, and practices can bring about a fresh awareness of ourselves. By looking at our teaching selves, we discover things about our personal selves. In other words, we can gain greater insight into who we are as people, including what we like about ourselves and what we don't like. However, such introspective awareness does take a bit of bravery. As Jersild (1955) points out in *When Teachers Face Themselves*, "to gain in knowledge of self, one must have the courage to seek it and the humility to accept what one may find" (p. 83).

Development also means that teachers can liberate themselves from prescriptions about what and how to teach. Teachers lose the need for prescriptions about how they should teach from supervisors and teacher educators. I have found that once teachers gain a commitment to and the skills to study and reflect on their own teaching and to know how to make their own informed teaching decisions, they do not always appreciate prescriptive advice about how they should be teaching. In short, the benefit is being attitudinally liberated from the prescriptive confining voices of those who believe they know the best ways to teach. However, the drawback is that having this attitude sometimes makes doing our job difficult, especially if those in power insist that those in subordinate positions follow their prescriptions without question.

But, to me personally, one of the greatest benefits to the teacher is that development is empowering. Teachers who consistently work on developing their teaching become empowered to make their own informed teaching decisions through awareness of their own knowledge, skills, beliefs, and attitudes. And, being able to make more informed teaching decisions can benefit the students. We cannot guarantee that students will learn more language as a direct result of development, but, certainly it is logical to believe that students have a better chance of learning when they study with an informed aware teacher who is able to make informed teaching decisions.

It is easy to define teacher development, point out aspects of teaching that can be developed, describe and illustrate teacher development activities, and point out benefits of teacher development. However, it is far more difficult to understand the complexity of teacher development in Korea. This is something that administrators, teacher educators, and language teachers in Korea need to consider. How much emphasis is placed on teacher development? How much emphasis is on prescriptive training? Is there a need to focus attention on teacher development? If so, what kinds of teacher development activities will meet the specific needs and interests of language teachers in Korea? What ramifications does an emphasis on teacher development have for preservice and inservice teacher education? Are administrators and teacher educators willing to further empower teachers to be able to make their own informed teaching decisions? Are teachers willing to take on this responsibility?

The Author

Dr. Jerry G Gebhard is a professor at Indiana University of Pennsylvania, United States, where he teaches in the Graduate program in Rhetoric & Linguistics. he has published widely in the area of English Teacher Education and Development.

References

Akamine, M. (1991). *An investigation of vocabulary teaching*. Unpublished course project. MA TESOL Program, English Department, Indiana University of Pennsylvania.

Bailey, K. M. (1990). The use of diary studies in teacher education programs. In J. C. Richards & D. Nunan (Eds.), *Second Language Teacher Education* (pp. 43-61). New York: Cambridge University Press.

Bailey, K. M., et. al. (1996). The language learner's autobiography: Examining the "apprenticeship of observation". In D. Freeman & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp. 11-29). New York: Cambridge University Press.

Brock, M. N., Yu. B., & Wong, M. (1991). "Journaling" together: Collaborative diary-keeping and teacher development. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Perspectives on second language teacher education* (pp. 295-307). Kowloon, Hong Kong: City Polytechnic of Hong Kong.

Carr, W. & Kemmis, S. (1986). *Becoming critical*. London: The Falmer Press.

Clarke, M. A. (1982). On bandwagons, tyranny, and common sense. *TESOL Quarterly 16*, 437-48.

Crookes, G. (1993). Action research for second language teachers: Going beyond teacher research. *Applied Linguistics 14*(2), 130-44.

Edge, J. (1992). *Cooperate development*. Essex: Longman.

Ellis, R. (1986). Activities and procedures for teacher preparation. *ELT Journal 40*, 91-99.

Fanselow, J. F. (1977). Beyond Rashomon: conceptualizing and observing the teaching act. *TESOL Quarterly 11*, 17-41.

Fanselow, J. F. (1987). *Breaking rules: generating and exploring alternatives in language teaching*. White Plains, NY: Longman.

Fanselow, J. F. (1988). "Let's see": contrasting conversations about teaching. *TESOL Quarterly 22*(1), 113-30.

Fanselow, J. F. (1992). *Contrasting conversations*. White Plains, NY: Longman.

Fanselow, J. F. & Light, R. (Eds.). (1977). *Bilingual, ESOL and foreign language teacher preparation: Models, practices, issues*. Arlington, VA: TESOL.

Farrell, T. S. C. (1996). *A qualitative study of four experienced EFL teachers as they reflect on their work*. Unpublished Ph.D. dissertation, Indiana University of Pennsylvania.

Flowerdew, J. Brock, M. & Hsia, S. (Eds.). (1992). *Perspectives on second language teacher education*. Kowloon, Hong Kong: City Polytechnic of Hong Kong.

Freeman, D. (1989). Teacher training, development and decision-making. *TESOL Quarterly 23*, 27-45.

Gebhard, J. G. (1991). Seeing teaching differently: The teacher as observer. *The Language Teacher 15*(5), 17-20.

Gebhard, J. G. (1993). The materials/media fair. In D. Freeman & S. Cornwell (Eds.), *New ways in teacher education* (pp. 56-59). Alexandria, VA: TESOL.

Gebhard, J. G. (1996). *Teaching English as a foreign or second language: A teacher self-development and methodology guide*. (Chap. 2) Ann Arbor: The University of Michigan Press.

Gebhard, J. G. (in press). *Awareness of language teaching: A guide to exploring and developing beliefs and practices*. New York: Cambridge University Press.

Gebhard, J. G. & Ueda-Motonaga, A. (1992). The power of observation: "make a wish, make a dream, imagine all the possibilities!". In D. Nunan (Ed.), *Collaborative language learning and teaching*. Cambridge: Cambridge University Press.

Jersild, A. T. (1955). *When teachers face themselves*. New York: Teachers College Press.

Johnson, K. (in press). *Teachers pedagogical thinking*. Boston, MA: Heinle and Heinle.

Jorden, E. H. & Noda, M. (1987). *Japanese: the spoken language*. Hartford: Yale University Press.

Kemmis, S. & McTaggart, R. (1982). *The action research planner*. Victoria, Australia: Deakin University Press.

Lange, D. L. (1990). A blueprint for a teacher development program. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 245-268). New York: Cambridge University Press.

Larsen-Freeman, D. (1983). Training teachers or educating a teacher. In J. E. Alatis, H. H. Stern, & P. Strevens (Eds.), *GURT '83 - Applied linguistics and the preparation of second language teachers: Toward a rationale* (pp. 264-274). Washington, DC: Georgetown University Press.

Lewin, K. (1948). *Resolving social conflicts*. New York: Harper and Brothers.

Li, D., Mahoney, D., & Richards, J. C. (1994). *Exploring second language teacher Development*. Kowloon, Hong Kong: City Polytechnic of Hong Kong.

Pennington, M. C. (1990). A professional development focus for the language teaching practicum. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp.132-151). New York: Cambridge University Press.

Richards, J. C. (1987). The dilemma of teacher education in TESOL. *TESOL Quarterly 21*, 209-226.

Richards, J. C. (1990). *The language teaching matrix*. New York: Cambridge University Press.

Richards, J. C. & Lockhart, C. (1994). *Reflective teaching in second language Classrooms*. New York: Cambridge University Press.

Richards, J. C. & Nunan, D. (Eds.).(1990). *Second language teacher education*. New York: Cambridge University Press.

Strickland, D. S. (1988). The teacher as researcher: toward the extended professional. *Language Arts 65,* 754-64.

Wallace, M. (1996). Structured reflection: The role of the professional project in training ESL teachers. In D. Freeman & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp. 281-294). New York: Cambridge University Press.

The Accuracy Of Interview Data: Testing Oller

MATT PEACOCK
City University of Hong Kong

This article describes an investigation into the correlation between learner answers on post-class materials evaluation questionnaires, filled out anonymously, and comments made by the same learners during subsequent interviews. The aim of the investigation was to determine the extent of learner truthfulness during the interviews, and thereby to make a preliminary assessment of the correctness of the suggestions of Oller (1978, 1981, 1982), Gardner (1980), and Bacon and Finnemann (1990, 1992) that learner responses to self-report data-collection instruments may be affected by self-flattery, responding in a way they are expected to respond, or responding in a socially acceptable way.

Data were collected daily over the entire seven-week 1994 summer term in two Yonsei University EFL classrooms, as part of a wider project researching classroom materials. Results indicated that the mean agreement between the questionnaire response and interview response for individual learners was fairly high at 69 %.

INTRODUCTION: LITERATURE REVIEW

There appears to have been little investigation into the accuracy of self-report data collected from adult EFL learners via interview and questionnaire. The accuracy of these data has, however, been called into question: Oller (1981) suggests (also refer Oller and Perkins, 1978a; Oller and Perkins, 1978b; and Gardner, 1980) that learner self-report data, particularly on their attitudes or motivation, are likely to be influenced by self-flattery, and / or a desire to be socially acceptable, and / or a desire to be consistent with their own previous statements. Oller (1981) continues by asserting that data-collection instruments in which these variables operate are weak in validity and can hardly be used to test theories empirically. He adds in a later article (Oller, 1982) that the self-flattery factor alone may result in up to 25 % variance in results (it should be noted, however, that while other researchers agree that caution should be used when drawing conclusions from learner self-report data, they do not go as far as Oller in questioning the value of the data). Bacon and Finnemann (1990) suggest that learners "may respond in a way that they believe they are expected to respond"(p. 461), and later (Bacon & Finnemann, 1992) propose it is possible that learners answer self-report questionnaires not "in the way they really believe", but rather "in what they perceive to be a socially appropriate way" (p. 491).

There have also been repeated calls in the literature for continuing work in this area. Oller and Perkins (1978b) describe it as "imperative" that researchers in the field "put up incontrovertible evidence that they have devised more valid measures...we do not believe that the validity of commonly used measures has been adequately shown"(p. 422). Oller (1981) observes that "If the measures of affective variables lack validity, they can scarcely be used as a basis for empirically testing theories"(p. 232). In a later article, he states that the "possible lack of validity" of such

measures "needs to be considered and, hopefully, ruled out by appropriate research" (Oller, 1982 p. 189). Bacon and Finnemann (1990) ask "can a self-report instrument designed to elicit learner beliefs about...affective input, in particular, reach an acceptable level of internal consistency?"(pp. 461-462) They argue that learner responses should be used to predict future learner responses, and also call for "additional research"(p. 466) focusing on the topic. Finally, Bacon and Finnemann (1992) call for "additional observational and experimental research"(p. 491) on learner responses to self-report measures.

As justification for the present study, first, I put forward the above indications of interest in the topic and calls for further research. Second, I suggest that the importance of learner interviews as a method of data collection for research on second language acquisition, particularly as back-up data designed to illuminate and explain results obtained from quantitative data, warrants investigation into learner truthfulness during interviews. A measure of how much reliance can be put on interview responses, particularly on the topic of motivation, certainly could have implications for calibrating the relative value of interview data.

RESEARCH METHOD

NOTE: The questionnaire and interview data on which this article reports were collected during a larger project designed to investigate whether authentic materials affect the classroom motivation of adult EFL learners. This topic was the focus of all questionnaires and interviews. The present paper reports only on the correlation between the questionnaire and interview responses of individual learners. For details of the larger project, including a copy of the questionnaire, refer to Peacock (1997).

This section is divided into five subsections. The first describes the two data-collection instruments used; the second, the subjects who participated in the study; the third, collecting the data; the fourth, the evaluation of the reliability of the questionnaire undertaken after the study; and the fifth, the evaluation and analysis of the interview data.

Data-Collection Instruments
(i) The Student Post-Class Questionnaire

As no existing questionnaires suitable for the study could be found, one from Gliksman, Gardner and Smythe (1982) was adapted. It was highly structured and was filled out anonymously by individual learners. The aim of this written questionnaire was to collect data on learner classroom motivation-specifically, to allow learners to express their opinion on the motivational value of the supplementary materials they had worked with on the day in question.

The questionnaire was translated for this project into Korean, to avoid learner misunderstanding: the English and the Korean translation were given together. Among the directions printed at the top of the form was one telling learners that the function of the form was to gather their opinions on the materials used. The questionnaire consisted of a list of seven items in opposition. Multiple items, chosen for their lexical similarity, were chosen in order to increase the reliability of the instrument. The items were: interesting / boring, unenjoyable / enjoyable, meaningless / meaningful, dull / exciting, satisfying / unsatisfying, unappealing / appealing, and absorbing / monotonous. Items were graded on a seven-point semantic differential scale. Learners put an 'X' against the place on the scale that best described their opinion of the materials. For example, an 'X' next to 'boring' scored one on that item; an 'X' next to 'interesting' scored seven; and an 'X' halfway between the two words scored four. A total score of between 7 and 49 was thus produced.

The purpose of correlating individual learner questionnaire responses with their own interview responses was to evaluate the level of learner truthfulness during those interviews. I suggest that the relative value of the questionnaire data collected in this study for the purpose of this correlation was

increased by three factors.

(1) The questionnaires were filled out anonymously by learners. I propose that responses on self-report questionnaires filled out anonymously are clearly much less likely to be affected by self-flattery, because learners would know that their names would not be connected with their responses. The same is true for the desire of learners to be socially acceptable, consistent with previous statements, and/or to respond in a way they are expected to respond.

(2) The questionnaire was piloted and revised until it reached a high level of reliability (this will be discussed in subsection four, 'Evaluating the Questionnaire'). It is a contention of this study that high levels of questionnaire reliability and consistency increase the reliability of correlational results obtained using data from that questionnaire — for example, correlation between questionnaire and interview responses.

(3) That the questionnaire was translated into the L1 of the learners very much reduces the possibility that it was a "surreptitious measure...of language proficiency" (Oller & Perkins, 1978a p. 85) of the learners, rather than a measure of their opinions or attitudes.

Although the questionnaires were completed anonymously, it was necessary, in order to allow the correlation of individual questionnaire and interview responses to be done after the study, for this investigator to write (after class) a number on each questionnaire identifying which learner had completed it. This was done by collecting the forms clockwise around the class every day, and by using a seating plan. It was done without the knowledge of the learners — had the learners known, the advantages of using questionnaires that had to be filled out anonymously would have been lost. However, I propose that the potential results justify the use of this technique, and that the technique is well within the guidelines for ethics for classroom research put forward by Hopkins (1985), and by Seliger and Shohamy (1989).

(ii) The Learner Interview

The aim of the semi-structured interview was to allow learners to give in their own words their views on the materials they had worked with on the day in question.

An interview sheet was used to question learners one-by-one after class in a five-minute interview. All interviews in both classes were done by this investigator. Learners were asked one question — "What is your impression of these materials compared to other materials used in this class so far?" The materials were shown to the learner as the question was asked. Answers were recorded in note form. There was a space on the sheet for student number, to enable the correlation between questionnaire scores and interview comments to be checked for individual learners after the study. Two learners were interviewed on a rotating basis on each day of the study. All learners were thus covered equally over the term; interviewees were not volunteers. Only two learners were interviewed each day because interviews took around five minutes each, and it was deemed unfair to ask the rest of the class to wait for their turn every day.

It was anticipated that some learners might be reluctant to criticise openly materials during interviews, particularly face-to-face with their own teacher (teachers seem to be regarded by some or many Korean learners as having a higher social status).

Subjects

Two upper beginning EFL classes took part in the study, containing 31 learners in all (16 in one class, 15 in the other) — 18 male and 13 female. The average age was 20, ranging from 18 to 24. All the learners were university students in Yonsei University's EFL institute in Seoul.

Collecting the Data

The research took place in two EFL classes assigned to this investigator for the 1994 summer

term. Each class ran for two-and-a-half hours a day for seven weeks. Supplementary materials used were a mixture of problems, task descriptions, questionnaires, puzzles, songs, and picture stories.

Questionnaires were collected every day for seven weeks (over 20 separate days, as class was held four days a week, interrupted by holidays). The stages of all activities were the same — the teacher introduced the activity, handed out the materials, went over the task, divided the class into groups of three, and started the activity. When the whole class had finished working and feedback had been conducted, learners filled out the questionnaires. A questionnaire was completed by each learner daily and then handed back to the teacher. Learners were regularly reminded that the focus of the questionnaire was the materials they had used on the day in question, and not the activity, their performance, or their teacher's performance. Questionnaire response rate was 100 percent, as learners filled them out in class every day. A total of 516 questionnaires was collected over the term.

Two learners were interviewed at the end of each of the 40 data-collection sessions (20 with each class), making a total of 80 interviews conducted over the seven weeks of the study. As there were 80 interviews, only 80 out of the 516 questionnaires collected - those completed by that day's interviewees - were required for the correlation on which this article reports.

Evaluating the Questionnaire

The reliability of the post-class questionnaire was evaluated via item analysis, factor analysis, and an alpha reliability check:

(1) The item analysis carried out on the questionnaires was based on all 516 questionnaires collected. Significance levels of $p < .001$ were computed for correlations between all items, indicating adequate learner comprehension of the meaning of all items.

(2) The aim of the factor analysis carried out on the 516 questionnaires collected was to check the relationship among the seven items on the form — that the items were really related to each other, making up one 'motivational value' factor. Results confirmed that this was the case.

(3) The overall alpha reliability for the questionnaire was computed at $r = .9103$ ($p < .0001$, $N = 516$). A day-by-day alpha reliability check was also carried out — that is, reliability was individually checked for all 20 days of the study, to ascertain whether learners continued to fill out the questionnaires with care throughout the study or not. Consistently high levels of reliability were found; alpha reliability never dropped below $r = .83$.

I suggest that results from the overall alpha reliability check on the questionnaires, as well as from the day-by-day alpha check, indicate a consistently high level of reliability for the instrument.

Evaluating and Analysing the Interview Data

One problem that arose with interviews was that they were conducted in English. This was done, first, because this investigator knew that learners would be less likely to object to giving up their own time after class occasionally if they saw those interviews as further English practice, speaking with their teacher, even if on a topic nominated by the teacher — classroom materials; and second, because it was possible during the interviews to ask for clarification when necessary, which was not possible with the daily post-class questionnaires. Unfortunately, though, conducting the interviews in English did have the effect of restricting the responses of these beginning-level learners to some extent. In retrospect, I decided that it would have been better to have spoken Korean with the learners (to clarify answers, at least) in order to elicit more accurately learner feelings and opinions about the materials.

Responses varied greatly and, in sum, comprise a mass of useful and interesting learner comments on and impressions of the materials brought to class (a number of the comments are quoted in Peacock, 1997).

Analysis of interview data was accomplished after the study in the following stages:

(1) each of the 80 responses was coded "positive" or "negative";

(2) responses were totalled;

(3) intra-rater reliability of the analysis was assessed, as suggested by Seliger and Shohamy (1989), by re-rating half of the comments some time after the original coding. This was done one month later. A very high level of agreement was obtained, indicating an acceptable level of reliability for the procedure;

(4) the overall correlation between interview responses and questionnaire scores was calculated (the results of this correlation will be reported in the next section, 'Results').

It was noted that learners were not afraid to speak out against materials — 33 out of the 80 comments were negative. However, the following problems arose with interviews:

(1) It was noticed during the interviews that some learners were being less than completely truthful, perhaps because they did not consider themselves equal in social status to the teacher, and thus did not want to make negative comments. This factor reduces the value of interview data. It may be more common in Korea than in the West.

(2) Interviews were conducted in English, meaning that (as discussed above) learners, at times, could not say precisely what they wished to say.

(3) It was only possible to interview two learners every day (as related above). This meant that the comments of other learners in class that day were necessarily lost.

An assessment of the relative accuracy, and therefore usefulness, of interview data was difficult to make until after the comments of individual learners had been correlated with their questionnaire scores. This will be discussed in the next section.

RESULTS

After the study, the correlation between individual learner interview responses and their own post-class questionnaire responses was checked in the following stages:

(1) All interview responses (which had been coded "positive" or "negative") were compared with the questionnaire score of the learner in question for the same day (for the purposes of this correlation only, these questionnaire scores were also coded "positive" or "negative", according to whether they were above or below that learner's mean questionnaire score for the whole study, averaged with the class mean questionnaire score for the day in question). For example, if a learner made comments coded as "positive" during the interview, and that learner's questionnaire score was found to be above the mean score as described above, this would be classified as "agreement" for the purposes of this correlation.

(2) Agreements and disagreements were totalled and the numbers compared. Results from number (2) above were as follows:

Agreement - 69% ($N = 55$).

Disagreement - 31 % ($N = 25$).

DISCUSSION AND CONCLUSIONS

It is proposed that the figure of 69% agreement is fairly high, and that it is some indication of a reasonably high level of learner truthfulness in interviews during the present study. I also propose that the strength of this indication is increased by the reliability of questionnaire data; learners could have had little motivation to be untruthful on questionnaires, as they were completed anonymously. Although it is possible that learners could have altered their interview comments to fit with the score they had just given on the questionnaire, I suggest that this probably did not occur.

The present study attempted to calibrate the relative truthfulness, and therefore value, of the interview data obtained from the subject classes. The results enabled us to decide how much reliance could be placed on those data for the purposes of the larger study mentioned above. I have also

partially responded to the calls of Oller and others (reviewed above) for further research in the area.

However, the correlation method used was relatively inexact — for example, if any one interview comment was slightly "negative", but the questionnaire score very slightly above the mean, the correlation would have been classified as negative, even though it may have been very close. This inexactness is to some extent a result of the relatively unsophisticated coding method used in the study for the analysis of interview data, and also of the difficulty inherent in correlating quantitative and qualitative data.

I suggest that investigators undertaking further research on this topic should use an improved and more precise system for coding interview data. Further research might also consider topic as a variable in the study. The focus of the data-collection instruments used in the present study was learner motivation; different topics may well engender different levels of learner truthfulness. Another method of calibrating the relative usefulness of the methodology and results would be to set up a control group. Learners in the control group would have their names written on their questionnaires, while learners in the experimental group would complete them anonymously. It would be possible to test the null hypothesis that there would be no significant difference in levels of truthfulness between control and experimental groups. Finally, all learners should be interviewed daily, if possible, in order to increase the amount of data available for checking the correlation between data-collection instruments.

Only speculation is possible at present as to why the correlation found in the present study was at the level of 69%. I speculate that at least a small amount of the 31% disagreement indicated by results was a consequence of the inexactness of the coding system used; and that learners were, on the whole, truthful during interviews.

As final conclusions, I suggest that the present results, while very limited in scope, are of interest and value as an example of the level of learner truthfulness in interviews during a study — particularly because the written questionnaires, scores from which were correlated with interview responses, were filled out anonymously. Skilled interviewers researching second language acquisition may believe that they are obtaining a high level of truthfulness during interviews. The veracity of this belief should perhaps be checked, using empirical data collected in existing classrooms.

THE AUTHOR

Dr. Matt Peacock is an assistant professor in the English Department of City University of Hong Kong. Before that he taught at Yonsei University, Seoul Korea.

REFERENCES

Bacon, S. M., & Finnemann, M. D. (1990). A study of the attitudes, motives, and strategies of university foreign language students and their disposition to authentic oral and written input. *Modern Language Journal 74*(4), 459-473.

Bacon, S. M. C., & Finnemann, M. D. (1992). Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input. *Language Learning 42*(4), 471-495.

Gardner, R. C. (1980). On the validity of affective variables in second language acquisition: conceptual, contextual, and statistical considerations. *Language Learning 30*, 255-270.

Gliksman, L., Gardner, R. C., & Smythe, P. C. (1982). The role of the integrative motive on students' participation in the French classroom. *Canadian Modern Language Review 38*, 625-647.

Hopkins, D. (1985). *A Teacher's Guide to Classroom Research*. Milton Keynes: Open University Press.

Oller, J. W. Jr. (1981). Can affect be measured? *International Review of Applied Linguistics for*

Language Teaching 19(3), 227-235.

Oller, J. W. Jr. (1982). Gardner on affect: a reply to Gardner. *Language Learning 32*(1), 183-189.

Oller, J. W. Jr., & Perkins, K. (1978a). Intelligence and language proficiency as sources of variance in self-reported affective variables. *Language Learning 28* (1), 85-97.

Oller, J. W. Jr., & Perkins, K. (1978b). A further comment on language proficiency as a source of variance in certain affective measures. *Language Learning 28*(2), 417-423.

Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *English Language Teaching Journal 51*(2), 144-156.

Seliger, H. W., & Shohamy, E. (1989). *Second Language Research Methods.* Oxford: Oxford University Press.

Student Pronunciation: A Comparison of Evaluation Techniques

PETER E. NELSON
Chung-Ang University

Students in a beginning English pronunciation class were asked to read two literary passages onto cassette tapes, which were then analyzed by the instructor for syllable stress, vowel and consonant articulation, intonation, rhythm and naturalness. Scores on a 0 - 100 scale were then assigned; after one week the material was re-evaluated using the same scale. Correlation analysis then determined the degree of reliability for each indicator between the first and second reviews. The resulting pattern suggests an inverse relationship between specificity of the pronunciation unit and reliability measurement. Highest ($r = 0.7$) correlations were for naturalness and an aggregate variable, Sum, obtained by grouping the characteristics. Lowest correlations were for vowel and consonant articulation and intonation. It was also found that the most reliable oral examination method is for a teacher to review each tape twice then average the scores. These findings have widespread implication faaor pronunciation testing techniques.

The growing emphasis on communicative approaches for the teaching of English has placed higher demands for correct pronunciation on teachers and students alike. Teachers have had to become more aware of pronunciation patterns of their students, as well as ensuring their own speech is correctly pronounced. Students in turn have had to listen more and have had to spend more time learning conversation skills. In addition, in some countries (e.g. Korea) tests for both listening and speaking now comprise portions of high school and college entrance examinations.

Numerous exercises have been developed to assist teachers in teaching and evaluating pronunciation. Many of these exercises focus on individual components including phonemes, syllables and individual words, while others emphasize short phrases, role play, structured or free conversation. Accompanying evaluation techniques: cloze exercises, identifying correct sounds in minimal pairs, locating stress within words, checking for pronunciation and so on, also have been developed to provide feedback to students, highlighting their oral strengths and weaknesses. When used correctly they isolate the sounds of English and enable students to enunciate them properly.

Given this emphasis on pronunciation, a question that naturally arises concerns the degree to which teachers can effectively and consistently evaluate speaking characteristics of students in extensive connected speech, whether students are reading a passage or are engaged in conversation. Simply put, are teachers reliable[1] in their evaluation techniques; can they consistently and correctly discriminate among student speech patterns in order to give accurate grades? The answer is not an automatic 'yes' or even 'possibly,' and as such deserves systematic examination. Moreover, knowing the extent to which teachers can effectively judge their students is of practical concern, as professional educators as well as academics should know how much they can rely on graded speaking characteristics as indicators of student verbal achievement.

---

1) In a testing environment, a reliable test is "one that produces essentially the same results consistently on different occasions when the conditions of the test remain the same." (Madsen, 1983, p. 179) In terms of statistical analysis, it makes no difference whether a student performs the same test two or more times, or the teacher evaluates the same student performance two or more times.

This paper examines to what extent one teacher (the author) reliably and systematically evaluated major characteristics of connected speech. Its focus is not free or structured conversation, but student readings of two passages which offered a variety of phonotactic features and facilitated differentiation among them. As presented, then, the students' pronunciation characteristics should not be considered as proxies for their performance in conversation, since intonation patterns, elision, linking and other features normally associated with free speech tend to be different when students are reading. In effect the methodological setting is deliberately limited, focusing on a teacher's listening skills in an artificial context.

Theory and Methodology

This paper is exploratory in that it originated from an experiment by the author to determine whether his own listening skills were reliable. To do so it was first necessary to ask which speech components he could identify, then design an experiment that systematically evaluated them. His effort was not theory based but was centered on implementation; he wanted a technique that was practical administratively, had content validity (i.e. it appropriately measured what it set out to measure), was comprehensive, and was useful to both student and teacher in evaluating pronunciation. By extension, he also wanted a replicable technique, one that other teachers could use to assist their own students in a reliable manner.

An initial effort to examine recorded conversation of students provided both a context and impetus for the present study. The author found he was able to identify and evaluate the following voice characteristics: syllable stress, intonation, vowel and consonant articulation, and rhythm. In addition, he was able to assign numerical scores to utterances, thereby providing individual voice profiles as well as range qualities for the class as a whole. Moreover, when individual speech components were cross-correlated the resulting matrix suggested recognizable *speech patterns*, evidenced by high ($p < .001$) statistical interrelationships. Similar correlation patterns were repeated in second and even third evaluations, and implied that a teacher with some listening skills could reliably assign scores, and grades, to recorded speech. While this initial finding was important, the original study was not systematic or guided by theory. It also was deficient in that each student recorded different material, thereby eliminating systematic evaluation of identical recordings.

Subsequent efforts to develop a research design guided by linguistic theory produced mixed results, at least in the Korean context. Both a literature search and conversations with Korean linguists suggest that reliability studies of language teachers are few and inadequately developed. Lee (1991), in his review of testing reliability literature generally, distinguishes between inter-marker reliability and intra-marker reliability. The former consists of scores given by different assessors to the same testing material, whereas the latter concentrates on the same assessor giving marks on a particular response. Lee cites several international studies of both categories for oral testing and notes that high correlations (.6 to .96) are possible for broadly based indicators like grammar, vocabulary, pronunciation and fluency (Lee, Chap. 5, 8). While this current study does not parallel Lee's efforts to develop reliable oral testing indicators, his work is useful by suggesting that reliability testing for teachers has occurred and can produce satisfactory results. It is also important in suggesting that assessors must have adequate training in listening techniques in order to achieve high reliability.

Although testing literature pertaining to Korea is sparse, the parameters of *pronunciation generally* are well known, as are both theoretical and empirical fundamentals in language testing procedures. These include the pioneering study of Chomsky and Halle (1968) for sound patterns of English; notions of 'articulatory settings' and 'anchorage' as applied to Korean speakers (Cha, 1989); measurements and indices of learner development in the target language (Larsen-Freeman & Long, 1991); testing techniques for language learners (Cohen, 1991), and considerations in language testing (Bachman, 1990, Bachman & Palmer, 1996). While these works are important for understanding language acquisition and evaluative processes, they do not specifically address intra-reliability testing of teachers, which is the primary focus of the current study.

This research design focused on three interrelated questions pertaining to the use of pronunciation tapes for diagnostic testing purposes:

a) If a teacher listens to identical material a number of times, is he/she likely to evaluate it exactly the same way, or with noticeable variability?[2.]

When considering language evaluation, other considerations will apply that may affect how teachers judge discourse. For example, the degree of phonetic/linguistic training the teacher has had, or their different academic or geographical backgrounds, may affect his or her "interpretation" of what is heard. (From a private communication by Dr. Wyn Roberts, Department of Linguistics, Simon Fraser University, British Columbia, Canada, 18 May 1998).

b) Which speech components are most consistently evaluated the same way?[3]

c) What evaluation technique is best in terms of reliability/consistency?[4]

A test for congruence patterns requires both a conceptual framework and an empirical evaluation technique. The framework for this study is the belief that methods of evaluating listening cannot be arbitrary but must be purposely designed to ensure that teachers accurately and consistently identify those key components of speech which can be accurately and consistently identified and, in EFL/ESL situations, accurately and consistently evaluated. Without this, the whole notion of EFL/ESL language pedagogy as a general activity becomes almost vacuous because of its arbitrariness. It is expected, however, that some interpretative variability exists if evaluative, non-mechanical listening techniques are employed. While speech characteristics have been analyzed by machines with mixed results[5], the current study wanted to provide a practical listening technique for teachers.

The context is a beginning English Education pronunciation class at a university in Seoul. Since only one restricted sample is evaluated, the study admittedly suffers from low variability in the range of speech characteristics. It benefits, however, from tighter methodological rigor and measurement, and as such provides a solid basis for further academic exploration. At this early stage no attempt is made to generalize to larger populations of students, but solely to address the three questions above first with regard to one teacher, and by extension to others.

Five speech components are analyzed concurrently: syllable stress, vowel and consonant articulation, rhythm and intonation.[6]

A sixth variable, naturalness, is used as a summative indicator, whereby the listener concentrates on the entire speech pattern rather than specific aspects. Both parametric and non-parametric

---

2) There is an expectation that the teacher will not evaluate material in an identical fashion, especially if there are many items to consider and/or many students to evaluate. This difficulty exists across the entire educational spectrum, as it is unlikely that a teacher's emotional state, knowledge, or other factors will remain constant

.3) Reliability and consistency in this context are also related to a third component: the specificity of the phonetic unit(s) being evaluated. "The more specific/finer the phonetic unit, the greater the likelihood of variability in noting, recording and interpretation by teachers… From a linguistic/phonetic theoretic point of view, there is a scale of smaller to larger and vice versa, and a corresponding inversely related scale of less-to-greater evaluation reliability." (Ibid.)

4) There are of course numerous ways to evaluate utterances, ranging from impressionistic and qualitative to mechanically based systems. In this paper the aim was to assist the teacher, to determine ways that are feasible and can be implemented .

5) "Machine-based analyses are not that successful, partly because they identify physical properties which still have to be interpreted as to their linguistic functioning. For example, pitch is not always in one-to-one correspondence with tone, amplitude with pitch, etc. If you have, for example (as in some ESL programmes) pitch extractor records of Chinese (e.g Mandarin) "tones", imitation is still not likely to be improved because the record is of one person, with a particular voice, etc. and other non-phonemic properties of consonants and vowels related to tonal variation have not been fully identified and properly interpreted." (Ibid.)

6) This study recognizes that recognition of stress placement, vowel articulation and so on are difficult and as concepts are themselves subject to interpretation. Nevertheless, these phonetic features are identified in pronunciation textbooks and are taught to students, so teachers presumably would want to test their students to ensure they pronounce "correctly." In addition, the study is examining one teacher only, not comparing listening skills of different teachers.

statistical tests are employed (Pearson product moment and Spearman rank order correlations, respectively). Definitions for the above terms are listed in the appendix.

IMPLEMENTATION

The study evaluated two separate listenings of 45 cassette tapes recorded by students in the class. Each student read the same two passages (see appendix), from which the six speech variables/characteristics were extrapolated and studied by the reviewer (myself). Tight measurement controls were maintained to ensure that first and second listening sessions did not overlap or influence each other. All identification of individual students was eliminated, and tapes were thoroughly shuffled, pooled, and drawn from on a random basis. The reviewer then listened to them all, briefly, to obtain background information necessary to compare subjects and to develop a scoring system.

A three-stage review process was used in order to test reliability. This included:

a) An initial review listening only for 'naturalness' of conversation, not for particular speech characteristics. A scale of 1-100 was used, using cut-off points of 0-59, 60-69, 70-79, 80-89 and 90-100. In classroom marking these ranges would often be classified as Excellent, Good, Adequate, Weak and Poor, respectively.
b) A review of the specific speech components using the same 0-100 scale. These were selected after consultation with a phonetics specialist.[7]
c) A repeat of the above steps using the same sequence and techniques.

Separate scoring sheets were used for the first and second listenings. At no time were scores compared between listening sessions, nor were scores adjusted later. The second listening was two weeks after the first.

Hypothesis testing

After tapes were scored, several methods were used to check for congruence patterns:

a) Each speech component was examined statistically for distribution including skewness, kurtosis, standard deviation, variance and standard error
b) A histogram was produced for each criterion
c) Both Pearson product-moment and Spearman rank order correlations were undertaken
d) An aggregative variable, Sum, was used to combine the separate indices of stress, vowel and consonant articulation, rhythm and intonation[8]
e) Reliability tests were performed on all interval data and transformed variables.

The first relationship to be tested concerned the separate speech components. The working hypothesis was that a teacher listening to the same material twice would grade it in the same manner, giving each student approximately the same score on the second evaluation as on the first. A high positive correlation was hypothesized for all interrelationships.[9] Following Cohen's(1991, p.496) observation on reliability coefficients, a correlation of 0.7 was selected as a 'meaningful' cutoff, to

---

7) The reviewer initially selected these characteristics from a well known pronunciation textbook used for our freshmen students; Judy Gilbert's (1993) *Clear Speech: Pronunciation and Listening Comprehension in North American English (2nd Ed.*) published by Cambridge University Press. He then showed the list to Dr. Hee San Koo, a Korean phonetics specialist at Chung-Ang University in Seoul, and to Dr. Kyung-whan Cha, a listening expert also at Chung-Ang. Neither Dr. Koo nor Dr. Cha, however, listened to the tapes themselves.

8) "Sum" is an artificial variable in that it does not measure a single component of speech. It should not be compared to "Naturalness," which is a personal assessment of a subject's overall speech pattern. It was devised and included in order to examine whether all individual indicators, when combined, gave a greater reliability score than any one indicator considered separately.

9) As noted by Madsen(op cit), a perfect correlation between first and subsequent evaluations of students' linguistic skills is not expected given human variability. The question is therefore framed in a probabilistic sense: is there a high likelihood that first and subsequent scorings will be similar?

indicate good reliability.[10]

The results are shown below both for product-moment and rank order correlations. In these and other tables the number "1" refers to the first listening and "2" to the second. Correlations above .7 are indicated in **bold type.**

Analysis of Tables 1 and 2 indicate that while all correlations are positive and significant statistically at the .01 level (one-tail test), the hypothesis of congruence itself is only partially upheld. Only the Naturalness indicator and the Sum indicator suggest acceptable congruence. In practical terms these unexpectedly low correlations suggest that reliability for these indicators between first and second listenings in not especially high.

This is an important observation.[11] If this pattern is commonly found in subsequent studies it

Table 1: Speech Component Interrelationships						
(Pearson Product Moment Correlations)						
	Natural1	Stress1	Vowel1	Conson.1	Rhythm1	Intonation1	Sum1
Natural2	**.755**						
Stress2		.642					
Vowel2			.579				
Conson.2				.591			
Rhythm2					.554		
Inton.2						.439	
Sum2							**.772**

Table 2: Speech Component Interrelationships						
(Spearman Rank Order Correlations)						
	Natural1	Stress1	Vowel1	Conson.1	Rhythm1	Intonation1	Sum1
Natural2	**.745**						
Stress2		.634					
Vowel2			.600				
Conson.2				.528			
Rhythm2					.468		
Inton.2						.378	
Sum2							.697

---

10) Cohen's cutoff of 0.7 is arbitrary and should be viewed in context. A correlation coefficient when squared gives the variance. In this case 0.7 when squared gives 0.49, or approximately half of the explained variability. Whether this amount is "reliable" is best left to the interpreter and to the situation under review.

11) This pattern corresponds at least to another professor's experience and observation — that the working hypothesis is not strongly borne out. (Roberts, op cit).

Student Pronunciation: A Comparison of Evaluation Techniques

may support the hypothesis of an inverse relationship between specificity of the phonetic unit studied and the degree of evaluation reliability.

A secondary observation concerned the relationship of interval and rank-ordered data. In each instance the correlation for the interval datum is higher than its ranked counterpart, although the degree of difference is relatively small. In practical terms the findings suggest that a teacher can evaluate students either by whole numbers or by ranks. The former method, however, provides greater opportunity to perform statistical tests, provided the data is normally distributed. All interval data variables used in this study had acceptable normalized distributions.

The second hypothesis concerned the strength of relationships among the separate speech components for each listening activity. The test here was whether correlations would be the same for each component of speech. As noted earlier when observing Tables 1 and 2, only the aggregative indices, Naturalness and Sum, correlated highly between first and second listenings.[12]

The implications of these correlations are substantial. They have bearing, for example, on such instances as students asking for specific details where marks have been deducted or grades otherwise adversely affected. The broad evaluation of parameters such as 'naturalness' makes this quite difficult— but it does get easier the larger the linguistic unit considered on (e.g. grammatical errors in phrases, clauses, sentences, etc.). However, pointing out errors in consonantal or vowel feature is not satisfactory or easy for the student (or teacher).

The above finding may also suggest a larger-to-smaller EFL/ESL language teaching process; concentrate on sentences (semantics/pragmatics, syntax, grammar first, together with/moving on to sentential rhythm and intonation), and finally on to the finer details of consonants and vowels. In terms of phonetic/linguistic theory, the direction of larger-to-smaller (the top-down / transformationalist / structuralist consitituent analysis) view is better than smaller-to-larger (bottom-up structuralist/Bloomfieldian) view.

Analysis of the remaining correlations suggests that Stress retains high congruence whether measured on an interval scale or a ranked one. Intonation is the lowest, with measures for Vowel and Consonant articulation, as well as Rhythm, having scores in the middle range. One conclusion drawn from the tables above suggests that teachers are likely to achieve a higher reliability score by listening for correct syllable stress rather than for 'finer' measures such as vowel/consonant articulation or intonation. These latter

Table 3: Rescaled Indicators						
(Pearson Product Moment Correlations)						
	Natural1	Stress1	Vowel1	Conson1	Rhythm1	Inton1	Sum1
Natural2	.657						
Stress2		.579					
Vowel2			.420				
Conson2				.367			
Rhythm2					.547		
Inton2						.385	
Sum2							**.737**

_____

12) The importance of these correlations in relation to problems associated with meaningful and reliable correction was indicated by Roberts (Ibid.).

measures appear elusive, at least in the ability to distinguish among 45 separate students in two separate listenings.[13]

Comparative evaluation techniques were employed to determine which evaluation method produced the highest indices of reliability and consistency. To do so required two data transformations:

a) Rescaling the 1-100 index to a simpler 1- 5 index for each speech component, then comparing the correlation matrices of scaled scores to original matrices. The reasoning behind this transformation is that judgment on a 1 - 100 index is far too difficult for a teacher to undertake, whereas giving a mark between 1 and 5 is easier and realistic. Accordingly, the scores were rescaled as follows: 5 = 90 - 100 ("Excellent"); 4 = 80- 89 ("Good"); 3 = 70 - 79 ("Adequate"); 2 = 60 - 69 ("Weak"); 1 = 1 - 59 ("Poor"). Results are shown in Tables 3 and 4.

Observation of the above tables shows that re-scaling the variables from a 1 - 100 basis to a 1- 5

Table 4: Rescaled Indicators						
(Spearman Rank Order Correlations)						
	Natural1	Stress1	Vowel1	Conson1	Rhythm1	Inton. 1	Sum1
Natural2	.630						
Stress2		.589					
Vowel2			.408				
Conson2				.330			
Rhythm2					.509		
Inton2						.346	
Sum2							.673

Table 5: Aggregative Scales: Pronunciation and Prosody								
(Pearson Product Moment Correlations)								
	Natural1	Stress1	Vowel1	Conson1	**Pron1**	Rhythm1	Inton1	**Pros1**	Sum1
Natural2	.755								
Stress2		.642							
Vowel2			.579						
Conson2				.591					
Pron2					.657				
Rhythm2						.554			
Inton2							.439		
Pros2								.533	
Sum2									**.772**

---

13) It should be strongly noted that the identification (by linguists/teachers) of rhythm and intonation is very unreliable. See, for example, Chomsky and Halle's (1968) discussion of stress levels in sound patterns of English.

marking system lowers congruence in all categories, and each item is lower than its interval-based counterpart. However, the relationships among the variables is retained, with Naturalness, Stress and Sum retaining the higher levels of congruence. This pattern also holds for rank order correlations, which show still lower levels of congruence for each category. In practical terms, re-scaling is less effective than using an interval scale with a greater range.

b) Combining scales. The reasoning here is that finely-tuned indicators such as Vowel or Consonant articulation, or Rhythm and Intonation, show low reliability between the first and second listening sessions. Collapsing two indicators into one aggregative variable would hopefully provide more accurate listening measures with higher reliability. Accordingly, vowel and consonant articulation were combined into an aggregative index called **Pronunciation**, while rhythm and intonation were combined into an aggregative index called **Prosody**. Results are shown in Tables 5 and 6.

The significance of the two tables above is the extent to which the process of combining variables achieves higher congruence than employing the non-aggregated indices independently. **Pronunciation** has higher correlation coefficients than either of its two sub-components(Vowel and Consonant articulation), while **Prosody** has higher correlations than either of its two sub-components(rhythm and intonation).[14]

Given the correlations above, another method of combining indices—by adding the first and second listenings to then obtain an average score—deserves review. Accordingly, a matrix was devised whereby all speech components were combined, including those of Pronunciation and Prosody. The findings are set out below.

The high correlations in Tables 7 and 8 suggest that a teacher attempting to judge students on pronunciation characteristics should undertake at least two tests, then combine or average the scores. This procedure appears superior to a single listening and to simplified ranking procedures. Moreover, it is also useful when combined with the two aggregative variables Pronunciation and

Table 6: Aggregative Scales: Pronunciation and Prosody								
(Spearman Rank Order Correlations)								
	Natural1	Stress1	Vowel1	Conson1	**Pron1**	Rhythm1	Inton1	**Pros 1**	Sum1
Natural2	.745								
Stress2		.634							
Vowel2			.600						
Conson2				.528					
Pron2					.617				
Rhythm2						.468			
Inton2							.378		
Pros2								.470	
Sum2									.697

---

14) This is to be expected in English since rhythm and intonation function together. Although the increase is small, in practical terms it *suggests that the teacher can test better for pronunciation and prosody than for the finer indices*. In addition, the combination creates time savings by testing for two indicators not four.

Prosody. The drawback, unfortunately, is that the procedure is quite time consuming, most likely limiting its usefulness in the classroom.[15]

Table 7: Combined Indices (Listening 1 + Listening 2)								
(Pearson Product Moment Correlations)								
	Nat	Stress	Vowel	Con	Rhythm	Inton	Pron	Pros	Sum
Nat1,2	---	.644	**.759**	**.717**	.583	**.702**	**.781**	.657	**.744**
Stress1,2			.786	.761	.754	.701	.819	.751	.876
Vowel1,2				.787	.798	.779	.948	.813	.911
Con1,2					.710	.771	.942	.760	.878
Rhy1,2						.886	.799	.975	.931
Inton1,2							.820	.967	.924
Pron1,2								.833	.947
Pros1,2									.955
Sum1,2									---

Table8: Combined Indices (Listening 1 + Listening 2)								
(Spearman Rank Order Correlations)								
	Nat	Stress	Vowel	Con	Rhy	Inton	Pron	Pros	Sum
Nat1,2	---	.568	**.705**	.650	.509	.648	**.725**	.586	.672
Stress1,2			.778	.746	.750	.700	.817	.754	.877
Vowel1,2				.768	.783	.741	.946	.78	.905
Con1,2					.667	.707	.927	.695	.841
Rhy1,2						.866	.780	.969	.920
Inton1,2							.773	.949	.888
Pron1,2								.797	.937
Pros1,2									.937
Sum1,2									---

15) Again, a larger issue is raised: to what extent is "usefulness" determined or affected by a teacher's clear and consistent ability to explain to a student where and what the errors are, why s/he is being marked ""wrong," and how to make improvements. Thus what may be useful to the student may be too cumbersome or time consuming to employ effectively in a large classroom. The above may be related to arguments over optimal class size for maximum efficiency/ optimal results in practical situations such as language teaching.

CONCLUSIONS

This study was designed as an initial step to measure how reliable teachers are when evaluating student pronunciation patterns on a repeat basis. A high degree of congruence among separate speech components was initially assumed, although it was expected that some variability would inevitably occur given the complex task of judging 45 students. The author tested his three working hypotheses using a number of statistical techniques and himself as the 'tested' teacher.

The major finding is that although high congruence levels do occur, the correlations are not uniformly high nor are the patterns consistent. In particular, greater variability—and less congruence—occurs the more finely-tuned the measurement indicators become. In contrast, aggregative indicators which combine a number of measures, or those that average repeated listenings, appear to be more normally distributed and show higher congruence levels. In statistical terms the aggregative indicators tend to 'smooth out' individual discrepancies in marking since they combine two variables. In practical terms, a teacher making a concerted effort to evaluate speaking characteristics would have to listen to each student's presentation two or more times, then average the scores.

Secondary findings relate to the type of measurement systems used. Interval scoring systems using a 1 - 100 scale give greater congruence than ranking systems, whether the ranks are done within a 1-100 system or on a simpler 1- 5 system. The differences are not great, however, and the convenience of a 1- 5 scale may outweigh the accuracy of a 1-100 scale.

There are several implications from these findings. One is that a teacher cannot expect uniformity or even high reliability in the examination of speech components, even though a variety of evaluative indicators is used. Too many factors intervene to permit mutual congruence, so teachers must employ several tests of pronunciation to develop a comprehensive profile of a class. In this regard the congruence pattern is roughly the same whether raw scores or ranking methods are employed.

Second, teachers who employ 'finely tuned' indicators (e.g. articulation, intonation) are likely to encounter greater error than if they use more aggregative indicators (e.g. 'naturalness', summary scores). For evaluation purposes, the finely tuned indicators may be employed to differentiate among students in order to provide specialized assistance, but the aggregative indices are better used for giving number grades.[16]

Third, having students read passages both helps and limits the teacher when assessing student speech characteristics. Identical readings known by students beforehand permit the teacher to evaluate closely how each student articulates words in context, and enables him to compare students using pronunciation markers. A drawback, however, is that student preparation for the reading(s) ultimately narrows the range of pronunciation differences among them.[17]

FUTURE RESEARCH

This paper is deliberately limited in that it has narrowed the choice of individuals to be studied, in this case to a beginning pronunciation class in an English Education department at a Korean university. Since the class was comprised primarily of students who wish to learn English in order to teach the subject, it might be expected that those studied would have many pronunciation characteristics in common. As a

---

16) The problem of explaining grades to students, and as well as ways to address this difficulty, however, remain unsolved. This deficiency may be covered in further research.

17) This is not an argument against prepared recordings or the reading of selected passages, since the reviewer still has a basis from which to compare subjects. Alternatively, prepared recordings or passages might be used by a reviewer to determine if there is a carry-over into free conversation, however literary the conversational style might be at first. In this regard the range of student responses is extended when they speak in open conversation, although the evaluation task faced by the teacher is more difficult. In this latter instance the teacher should trust his hearing skills and use the 'naturalness' indicator to distinguish among students, whether judging open conversation or their reading of passages.

consequence, their range of differences would be narrow, inhibiting a researcher to carefully distinguish among them. This observation is borne out by an earlier study undertaken by the author, in which the same students were tested using related indicators: the range of pronunciation skills was considerably greater at the beginning of the class than at the end. In other words, the class was successful in narrowing differences!

Future studies, therefore, should include research designs targeted for students at different age and education levels, as well as for broader segments of the population. Moreover, a battery of statistical techniques exists to analyze new studies. Multi-dimensional scaling, cluster analysis, multiple regression and factor analysis, for example, can provide a wealth of information on pronunciation patterns and related speech characteristics.

Efforts to re-design tests appropriate for Korean students must also be appropriate to the test providers. In this regard, Lee's (1991) analysis of testing within Korea suggests that Korean teachers of English are inadequately trained in listening skills. His suggestions for change include revised marking scales, broader testing categories and teacher training, all of which can improve both inter- and intra-reliability.[18]

Finally, this study has been limited to Korean students, whose speech patterns in English often differ substantially from those of EFL learners in other countries. Virtually unlimited opportunities exist for cross-national investigations, and the serious researcher should take advantage of them. The universal principle requires that the general issues, however, will remain the same everywhere.

THE AUTHOR

Peter E. Nelson, Ph.D., is an instructor in the Department of English Education at ChungAng University in Seoul, South Korea.

REFERENCES

Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.

Bachman, L F., & Palmer, A. S. (1996). *Language Testing in Practice*. New York: Oxford University Press.

Cha, K. W. (1998). Articulatory settings. *Journal of the Applied Linguistics Association of Korea 2*, 83-95.

Chomsky, N., & Halle, T. (1968). *The sound pattern of English*. New York: Harper and Row.

Cohen, A. D., (1991). Second language testing. In Marriane Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language*. Boston: Heinle and Heinle,

Gilbert, J. (1993). *Clear Speech: Pronunciation and Listening Comprehension in North American English (2nd Ed.)*. New York: Cambridge University Press.

Koo, H. S. (1997). A Study using Acoustic measurement and Perception to identify prosodic characteristics of English as spoken by Koreans. *Korean Journal of Speech Sciences 2*, 95 -

_____

18) Lee, op. cit., Ch. 5, 7, 8. The entire work relates to testing criteria. Although teacher training and oral examination techniques have improved since 1991, his analysis and recommendations are still relevant and useful.

108.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.

Lee, W. K. (1991). *A task-based approach to oral communication testing of English as a foreign language*. Seoul, South Korea: Hanshin Publishing.

Madsen, H. S. (1983). *Techniques in testing*. London: Oxford University Press.

<div align="center">

APPENDIX
DEFINITION OF TERMS:

</div>

- **syllable stress**. Stress is identified as greater force exerted in the articulation of one part of an utterance compared with another, thereby accentuating a certain part of the utterance to give it more prominence. Although it is possible to identify primary, secondary, and even tertiary stress within words, this study listened solely for correct placement of primary stress.

- **vowel articulation**. The study concentrated on both vowel length and clarity, especially on longer stressed syllables to identify primary stress.

- **consonant articulation**. The study concentrated on the clarity and correct pronunciation of consonants within words, whether as beginning sounds, endings or medials. In particular, it listened to articulation of stops and plosives, and to letters such as 'v' and 'f' that are difficult for Korean speakers of English to pronounce.

- **rhythm**. This is the pattern of stressed and unstressed syllables. The study listened for the pattern from the perspective of naturalness, i.e. is it the way a typical native speaker of English would read aloud or speak. No effort was made to time the rhythm pattern, but only to listen to it.

- **intonation**. This is the melodic pattern produced by pitch variations during speech. Although the study anticipated that intonation patterns for reading would perhaps sound more 'flat' than for conversation, it listened for intonation to signal contrast or identify focus words within the passages.

- **naturalness**. This was used as a summative indicator, in which the listener concentrates on the entire speech pattern rather than specific aspects like those noted above. It is also employed as a check, to see which separate components correlated most highly with 'natural' speech.

Reading Passage #1:

"The distribution and the use of English are changing, under two kinds of pressure: higher educational standards over the world as a whole mean that millions more people each decade become able to read and to understand English; the surging growth of communications by way of radio, television, films, newspapers, books, leads to more contact and especially more variety of contact with information and opinion, with science and literature, with people from other countries; and all these contacts are given expression in English. As these changes take place they bring individuals increasingly into contact with English not just as it is used in their own limited personal experience but in forms that are different from those they are accustomed to because they arise from societies, cultures, attitudes and usages other than one's own."

Reading Passage #2:

"It is a very long time, probably about three hundred years, since English was simply the language of the English. The language we know today as English developed by about 1400 out of the Anglo-Saxon dialects spoken in the southern part of England, with a good deal of borrowing from Norman French, the language of those who had conquered England in the year 1066. Wales, Scotland and Ireland were at that time separate countries whose inhabitants spoke Celtic languages."

Teaching Conversation Skills More Effectively

GRAEME CANE
National Institute of Education, Singapore

This paper has two main purposes: (a) to look at the ways in which conversation skills have been presented in EFL textbooks and courses, and (b) to suggest some alternative sources for and approaches to the teaching of conversation skills. The author argues that, despite the claims of linguistic success by the various EFL publishing companies, most textbooks do not provide language teachers with the information and materials they need to help English learners attain communicative competence. The paper suggests that a more direct, more linguistically aware approach to teaching the communicative strategies and formulas used naturally by native speakers is likely to provide more effective guidance for learners than the conventional indirect approach.

Wife: What watch?
Husband: Ten watch.
Wife: Such watch! *Casablanca (1942)*

INTRODUCTION

If you've seen *Casablanca*, you may remember the above conversation spoken by the elderly Eastern European couple in the film who spend time practicing their English together in preparation for life in the United States. As English language teachers, how would we react to this dialog? We know from our experience that learners like these will be faced with a huge range of conversation tasks when they eventually reach America. What can we do to help? We might act in a number of different ways; here are some suggestions:

1. We could argue that successful communication is already taking place in this conversation and that the errors are linguistically trivial.
2. We could show the couple a Rolex advertisement and ask them to practice telling the time in pairs.
3. We could ask them to read some written dialogs which involve telling the time.
???

This paper has two main purposes: (a) to look at the ways in which conversation skills are presented in EFL textbooks and courses, and (b) to suggest some alternative sources and approaches in teaching these skills.

The strong theoretical emphasis given to the 'communicative approach' in EFL methodology around the world over the last fifteen to twenty years has, surprisingly, had little more than a superficial effect on the ways in which EFL textbooks handle the acquisition of conversational competence. An examination of recent communicative-styled coursebooks shows that, while learners are regularly told to 'work in pairs and discuss the advantages and disadvantages of foreign travel, credit cards, eating meat', or to 'work in groups and discuss why people go on diets, cut down trees

or kill elephants', they are given remarkably little linguistic help about how to actually do these things.

Richards (1990) argues that communicative language teaching has generally taken an 'indirect approach' to the teaching of conversation skills. Advocates of the indirect approach believe that learner participation in interactive situations such as role plays, information gap exercises and problem-solving tasks will eventually lead to the attainment of communicative competence. Interactive situations are thus created for participation but little specific language input is provided. It is this approach, Richards believes, which communicative-based ELT textbooks have generally taken, and the textbooks examined by the author while preparing this paper support Richards' view.

TEXTBOOK APPROACHES TO TEACHING CONVERSATION

The 'Speaking' section of one well-known coursebook first asks students to ponder the question, "Are you a pessimist or an optimist?" and secondly asks them to select appropriate multiple choice answers to a short written questionnaire. The students are then told to "discuss (their answers) with other students". Up to the discussion point, this 'Speaking' section has thus been exclusively a reading exercise. Once the learners have answered the questionnaire, they are expected to make the jump from reading a wrriten text to talking about it in a coherent, natural way without any further help from the textbook writers. This expectation on the part of the authors is neither realistic nor honest, and the learner is generally left at the end of the exercise with little more than a feeling of frustration and failure.

Another text which was examined is just as unhelpful in its 'Speaking' sections. One unit, for instance, asks the learners to "put the following crimes in order depending on how serious you think they are" and then "discuss in groups". No further guidance about the kind of language a native speaker might use to discuss this topic is provided by the writers.

The authors of one book aimed exclusively at teaching conversation state in their introduction that it is "designed to give practice in idiomatic English conversation". Exercises include asking students to complete unfinished sentences in writing, to draw a map from their home to the center of town, to give written replies to questions such as "Is there a hospital in your community?" and to match written sentences with a set of pictures. In its 208 pages, the book does not present a single conversation between two or more participants. This text is thus a conversation book without conversations, which makes no attempt to show learners how English speakers go about the complex task of spoken communication.

As the above examples demonstrate, many EFL textbook writers today work on the assumption that speech is somehow naturally acquired once the learner has completed sufficient grammar exercises, listened enough times to the accompanying cassettes or has written sufficient answers to the reading comprehension questions provided. Such authors appear to believe that a printed text intended for visual author-reader written communication will help learners improve their speaking skills.

While conversation analysts have frequently noted the differences between speech and writing in terms of the linguistic forms and the discourse organization involved, textbook writers continue to use the medium of print to teach the medium of speech, refusing to acknowledge the possibility of a serious linguistic mismatch.

Marshall McLuhan highlighted this potential conflict by noting that spoken conversation involves multi-sensory communication (hearing, sight, touch, etc.), whereas writing involves a single sense (vision). McLuhan's often-quoted catchphrase, "The medium is the message", serves to warn us that the medium of communication influences the way in which a receiver interprets any given message. For McLuhan, speech-hearing is a 'hot' medium of communication because it is multi-sensory, whereas writing is a 'cool' single-sense medium (Miller, 1971, p. 104). According to

McLuhan, in spoken conversation, we are placed in a situation that calls many of our senses into play, and we use these senses to help us interpret a speaker's message. With a written text, however, the medium is exclusively visual and a completely different process of comprehension is at work. As a result, to try to represent multi-sensory speech on a single-sensory written page is likely to create serious communication problems. For the deconstructionist Jacques Derrida, the fundamental differences between speech and writing as systems of communication are based on a presence/absence opposition.

> Speech depends on spatial and temporal 'presence'. Speech and thought — nothing comes between them. No lapse of time, no surface, no gap. Writing (however) functions in 'absence'. The written marks are abandoned, cut off from the writer, yet they continue to produce effects beyond his presence and beyond the present actuality of his meaning. (Derrida, quoted in Collins & Mayblin, 1997, pp. 51-53).

Despite these warnings from McLuhan and Derrida, textbook writers have generally ignored the possibility of a media mismatch and have continued to use written texts to teach spoken conversation.

Some EFL textbooks are, admittedly, more helpful than the three teaching texts mentioned above but, considering all the insights into conversation provided by conversation analysts over the past few years, the techniques for developing conversation skills in language textbooks still remain remarkably crude. Constructed textbook dialogs, for example, generally bear very little resemblance to authentic native-speaker conversation because they exclude some of the most vital grammatical, pragmatic and sociolinguistic features of everyday spoken English. Contrast, for instance, the following fairly typical textbook dialog (Speakers A & B) with what actually happens in authentic native-speaker conversations (Speakers C & D, E & F, G & H).

A: Which sports do you like, John?
B: I like most sports. How about you?
A: I used to enjoy tennis, but now I like playing golf. Do you play golf too?
B: Yes I do. I'm going to play golf tomorrow. Would you like to join me?

C: Well what's the what's the failure with the football I mean this I don't really see I mean it cost the money how much does it cost to get in down the road now?
D: I think it probably it probably is the money for what you get you know erm I was reading in the paper this morning a chap he's a director of a big company in Birmingham. (Crystal & Davy, 1975, p. 19)

E: So I told her that I wouldn't let her borrow it again because the last time she -
F: - Oh, hi, Tony!
E: He's Sooo good looking!
F: Yeah, but he's going with Lisa Bradley. Anyhow, so you told her. (McLaughlin, 1984, p. 36)

G: I went this weekend to the pet store. That's my favorite stall in the whole mall.
H: Yeah, I like to go in there too.
G: So I went in- I went in there. We were lookin' to see if we could find - you know- just they probably wouldn't have. We were lookin' to see if we could find one. 'Cause I just think - my dad has one and I just think they're really neat dogs.
      (5 second pause)
H: They're real neat. But we don't have any pets, you know, we live in an apartment so it's kinda like. (McLaughlin, 1984, p. 192)

POSSIBLE CRITICISMS BY TEXTBOOK AUTHORS

In the face of the above comments, textbook authors and language course designers might wish

to argue that my assessment of the way conversation is handled in textbooks is erroneous for the following three reasons:

1. The textbook is designed to teach the linguistic system, i.e. to develop what Chomsky has termed 'competence'; it is not concerned with individual speaker 'performance'. As a result of this philosophy, textbooks present a series of decontextualised sentences as grammatical models to illustrate the function/structure/rule being taught. However, authentic language use always occurs within a specific context and must therefore represent 'performance'. To focus on competence rather than performance may teach the learner some interesting facts about the language system but not how to speak appropriately in real situations.

2. The textbook is intended for use with a classroom teacher who will use the book merely as a foundation on which to build a conversation skills program for his/her students. However, as we have seen, current EFL textbooks provide very little practical material for teachers to work with in the area of conversation development. Textbooks give us little useful information about, for example, the function of stress, the meaning of different intonation patterns, how to interrupt or close a conversation appropriately, and so on. If EFL textbook writers are serious about dealing with conversation skills development in a way that will be of genuine value to learners and teachers, they must try to discuss conversation features with the same clarity and explicitness they employ to discuss points of grammar.

3. The specially recorded audio cassettes which often accompany language courses are specifically aimed at developing spoken English. The underlying assumption is that, through listening to the accompanying tape, learners will acquire the linguistic and communicative competence they need to talk like native speakers. This assumption is, in general, inaccurate. Recorded tapes are usually designed to improve either pronunciation or listening comprehension and are not specifically aimed at developing the learner's productive speaking skills. *Task Listening* (Blundell & Stokes, 1981), for instance, contains a good selection of authentic spoken material. However, the tasks which the learner is given to perform after listening to the extracts are intended to demonstrate listening comprehension and usually involve a written rather than an oral response, such as identifying places on a map, filling in a form, etc.

ALTERNATIVE SOURCES FOR THE TEACHING OF CONVERSATION SKILLS
A direct approach towards teaching conversation skills

In their 1994 paper on teaching conversational skills, Dornyei and Thurrell note that second-language learners of English who have a good knowledge of English grammar and vocabulary may still fail as conversationalists because they have not acquired the appropriate rules or strategies involved in 'conversational competence'. Dornyei and Thurrell argue that the indirect approach to teaching conversation adopted by most communicative-style textbooks is less effective than a direct approach which provides explicit language strategies and input.

The direct approach involves designing a conversation program which will give the learner the specific microskills and strategies required for fluent conversation. Using this approach, Dornyei and Thurrell propose four categories which are largely based on the findings of recent work in conversation analysis. Included in the category of *Conversational Rules and Structures*, for example, are the following items: conversation openings, turn-taking, interrupting, topic shift, conversation closings, and so on. The authors suggest, with regard to topic shift or changing the subject, that the direct teaching of certain routine phrases such as, "Oh, by the way," "That reminds me of," "As I was saying," etc. would show learners how native speakers go about the somewhat tricky operation of changing the topic in a conversation.

Before closing a conversation, speakers typically use a sequence of pre-closing formulas as preparation for leaving. Dornyei and Thurrell argue that it is important to raise learner awareness

about how to end a conversation without sounding abrupt or rude. They suggest that closing strategies such as "It's been nice talking to you." "Well, I don't want to keep you." etc. should be explicitly taught.

In the *Social and Cultural Contexts* category, Dornyei and Thurrell propose the direct teaching of the relevant differences between formal and informal speech styles using a stylistic continuum of, for instance, "How do you do? / Nice to meet you / Hello / Hi / How are you doing?" etc.

Dornyei and Thurrell have incorporated their direct-approach methodology in a practical guide to teaching conversational skills entitled *Conversation and Dialogues in Action* (1992). The authors suggest using communicative activities to teach strategies such as how to start and keep informal conversations going, how to change the topic, how to use fillers to buy time in a conversation. They suggest a list of fillers (e.g. well, actually, you know, let's see, to be quite honest, etc.) and then develop communicative activities in which the fillers can be used appropriately (1992, p. 44). The Dornyei and Thurrell approach does not represent a definitive blueprint for teaching conversational competence, but their suggestions are useful in that they provide learners with practical guidance about how native speakers conduct conversations in English.

Interpreting implicatures

Bouton (1994) has noted that very few ESL textbooks deal with the interpretation of implicatures in conversation but, according to research he carried out in the United States, even reasonably proficient non-native speakers may interpret implicatures differently from native speakers of English. In the tests that Bouton conducted, the performance of the non-native speakers coincided with that of the native speakers taking the test on only five out of the total of twenty items. The following are examples of the questions used by Bouton in his implicature test.

Situation (a) : Two teachers are talking about a student's paper.

Teacher 1: Have you finished with Mark's term paper yet?
Teacher 2: Yes, I have. I read it last night.
Teacher 1: What did you think of it?
Teacher 2: Well, I thought it was well typed.

The test participants were then asked to choose one of the following options concerning Teacher 2's opinion of Mark's paper :

A. He thought it was good.
B. He thought it was important that the paper was well typed.
C. He really hadn't read it well enough to know.
D. He did not like it.

Native speakers would realize that Teacher 2's reference to the essay being 'well typed' seems to flout Grice's (1975) maxim of relation ("be relevant"), as the reply does not represent a satisfactory answer to Teacher 1's question. They would then look for the underlying implicature contained in Teacher 2's statement, which presumably conveys the meaning that the essay was poor in everything but its typing.

Situation (b) : Two roommates are talking about their plans for the summer.

Fran: My mother wants me to stay home for a while, so I can be there when our relatives come to visit us.
Joan: Do you have a lot of relatives?
Fran: Are there flies in the summer?

The test participants were then asked to choose the best interpretation of Fran's question:

A. Fran thinks her relatives are noisy.
B. Fran is new to the area and is trying to find out what the summers are like.
C. Fran has a lot of relatives.
D. Fran is trying to change the subject; she doesn't want to talk about her relatives.

Native speakers realise that the answer to "Are there flies in the summer?" must be "Yes" and that this should also be understood as the answer to Joan's question about Fran's relatives. If, as Bouton suggests, even advanced learners of English experience problems in interpreting implicatures, it would seem useful to incorporate specific exercises on implicature interpretation as part of the direct approach to teaching conversation advocated by Dornyei and Thurrell.

Providing direct language input in the conversation class

Pease and Garner (1985) claim that even native speakers need training and guidance in developing their conversation skills. They argue that many people need help in starting conversations effectively, in keeping the conversation going, in asking appropriate questions, and so on.

The guidance given by Pease and Gardner in their books and courses tends to be language specific. They provide explicit language examples which native speakers could adopt in real-life conversation situations. The following is an extract from their 1985 book;

If a woman introduces herself as a nurse, you could choose from these questions to ask:

"Why did you become a nurse?"
"What did you have to do to enter the field?"
"Tell me some of the problems that people often come to you with."
"How does listening to troubles all day affect your outlook on life?"

If you don't want to talk about her job, you could open-enddely ask;

"What do you do for fun when you're not nursing?" (Pease & Gardner, 1985, p. 33)

Whatever your feelings with regard to this type of conversation guidance for native speakers, it seems to me that the presentation of material like this in the E.F.L. classroom would provide English learners with useful language input for developing their conversational competence.

Using video and audio materials to provide direct input

If a language course were centered on audio/video material with the textbook taking on a subordinate role, our current preoccupation with using the medium of print to teach spoken communication might begin to decline. The video and audio material used would have to be appropriately contextualised to avoid the possibility of boredom previously experienced by many participants in oral drilling and pattern practice sessions. One of the major flaws of the audio-lingual method was that the spoken drills were generally based on decontextualised *written* language forms, rather than on natural spoken conversation. Video and audio material used during the conversation class should not be used to teach grammar points or listening comprehension, as is currently the case. Instead, the program would have to be specifically oriented towards productive speaking skills and genuine learner participation in, for example, spoken (not sung!) karaoke-style video dialogs. At the beginners' level, the audio tapes of Carolyn Graham's *(1979) Jazz Chants for Children* would seem useful material for providing learners with native-speaker conversation formulas using appropriate intonation patterns.

One area where the written textbook might still have a useful role to play in the teaching of speaking would be in supplying accurate written transcriptions of the audio/video material presented to learners. Crystal and Davy's *Advanced Conversational English* (1975) provided audio recordings of native-speaker conversations along with transcripts and comments on some of the paralinguistic features occurring in the conversation. Although the range of speakers and topics covered was rather narrow, Crystal and Davy's book/tape project represented an innovative step forward in its use of authentic spoken conversation rather than artificially constructed print-based dialogs. Unfortunately, their work has not been developed by mass-market EFL textbook writers.

CONCLUSION

Despite the constant claims of linguistic success and achievement by the various EFL publishing companies, it is simply not the case that textbooks have provided language teachers with the information and materials they need to help English learners speak more fluently, naturally and confidently in conversation. One reason for this would seem to be that textbook writers have generally ignored the linguistic mismatch involved in using the single-sensory visual medium of print to teach the multi-sensory medium of speech. The purpose of this paper has been to suggest that it is time to look elsewhere for guidance. This paper has proposed that a more direct, more linguistically-aware approach to teaching the communicative strategies and formulas used by native speakers in conversation would seem to be more helpful to learners than the conventional indirect approach. If we use the recent findings of conversation analysis to provide English learners with the language input and conversation strategies they will need, we may perhaps be able to set our students on a less frustrating and more rewarding route towards the attainment of conversational competence.

THE AUTHOR

Dr. Graeme Cane has taught at university level in South America, Saudi Arabia, Papua New Guinea, Brunei, Japan and Singapore. He has published on South-East Asian varieties of English, inter-cultural communication and language deviance. Dr. Cane is currently a lecturer in English language at the National Institute of Education, Nanyang Technological University, Singapore.

REFERENCES

Blundell, L., & Stokes, J. (1981). *Task Listening*. Cambridge: Cambridge University Press

Bouton, L. (1994). 'Conversational implicature in a second language: learned slowly when not deliberately taught', *Journal of Pragmatics 22*(2), 44-60.

Collins, J., & Mayblin, W. (1997). *Introducing Derrida*. New York: Totem.

Crystal, D., & Davy, D. (1975). *Advanced Conversational English*. Harlow, Essex: Longman.

Dornyei, Z., & Thurrell, S. (1992). *Conversation and Dialogues in Action*. Hemel Hempstead: Prentice Hall.

Dornyei, Z., & Thurrell, S. (1994). Teaching Conversation Skills Intensively. *ELT Journal 48*(1). Oxford: Oxford University Press.

Graham, C. (1979). *Jazz Chants for Children*. New York: Oxford University Press.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. Morgan (Eds.) *Syntax and Semantics Vol. 3*. New York: Academic Press Inc.

McLaughlin, M. (1984). *Conversation: How Talk is Organized*. Beverly Hills: Sage.

Miller, J. (1971). *McLuhan*. London: Fontana.

Pease, A., & Gardner, A. (1985). *Talk Language*. Singapore: Heineman.

Richards, J. C. (1990). Conversationally Speaking: Approaches to the Teaching of Conversation. In J. C. Richards (Ed.) *The Language Teaching Matrix*. Cambridge: Cambridge University Press.

The Oral Placement Test (OPT) Ten Years On: EFL Teachers constructing their own placement tests

THOMAS FARRELL
National Institute of Education, Singapore

Prior to 1986 the Language Center of Seoul's Yonsei University, placed students by written tests designed by non-native speakers, and the tests themselves were usually based on fill-in-the-blank type grammar exercises, with little relevance to communicating orally in English. In 1986, this author designed a standardized oral placement interview with a well-defined scoring system that would more accurately reflect the student's communicative ability and better ensure appropriate placement. At that time, results showed that the OPT had been successful in placing the students and that the drop out rate for students had fallen as a result. Now ten years later the same test has been studied for its validity and reliability. This paper briefly outlines the original study and the more recent study and then compares the results of both. Results show that the OPT, with a few minor changes, is still valid and reliable ten years on, thus supporting the notion that EFL teachers can construct their own tests that suit the context.

INTRODUCTION

Traditionally, placement in foreign language programs in non-English speaking countries has been determined by written tests designed by non-native speakers or commercial tests, and the tests themselves are usually based on fill-in-the-blank type grammar exercises, with little relevance to communicating orally in English. This was the method used at Yonsei University, Foreign Language Institute, Seoul, Korea up to 1986. As a result, after the start of each term, students had to scramble to change levels because they had been misplaced because of the written test. They either had superior conversational ability in English or they were placed at too high a level for them to follow and take part orally in the class. In addition to the instructor's complaints that the placement test lacked sufficient shared emphasis on grammar and communication, the test took over five hours to administer and grade, thus resulting in a high fatigue rate for the examiners. Also, this form of placement had conflicted with the goal of the Foreign Language Institute (FLI) which is primarily to increase oral language proficiency, with grammar and written improvement a secondary aim.

The purpose of the original study outlined in this paper was to develop a standardized oral placement interview with a well-defined scoring system that would more accurately reflect the student's communicative ability and better ensure appropriate placement. The greatest advantage of an interview for placement testing is that it provides an opportunity to simultaneously assess oral production (pronunciation and fluency) as well as the ease of language use in social situation (Harrison, 1983). This would now be in line with the above stated goals of the program offered to students.

The test developed (see Appendix A) was designed to be economical and practical in terms of cost, personnel requirements and time constraints for administering and scoring. The oral examination

had to allow for a large number of subjects who could be interviewed in as little time as possible, since most schools and institutions allow only one day for registration. This was one of the most important aspects of the OPT because as Clarke (1978) points out:

> No matter how highly valid and reliable a particular testing method may be, it cannot be servicable for 'real world' application unless it falls within acceptable limits of cost, manpower requirements, and time constraints for administration and scoring (p. 12).

This paper briefly outlines the test that was devised in the original study in 1986 (Farrell, 1986) and also outlines an updated unpublished study of the OPT ten years on (Gibb, 1996). The paper starts with a brief discussion of oral testing, then an outline of the study will be presented. Finally, an account of the update of the OPT is presented in the last section of the paper.

ORAL TESTING

Language tests are at present the only accurate method of indicating the English proficiency level of a student who wants to enter a language program. The two traditional methods of testing language proficiency levels of students have been written tests and oral interviews. Theorists and practitioners do not dispute the value of an oral interview for placement into these programs; rather, there are varying opinions as to what should be tested. A general proficiency test of interview indicates what an individual is capable of doing at a particular point in time as a result of his or her cumulative learning experiences (Harris, 1969).

What, then, is an oral English test? Upshur (1975) says that in an oral test, a tester first sets a task for the testee to perform. This task demands English speech after which the testee attempts to perform the task. Finally, the tester notes the performance of the testee and assigns a score. This definition sounds simple, but it is misleading as the assessment of spoken English has been a major problem for people involved in language teaching (Brown, 1983); it is more difficult to assess spoken English with exact accuracy than other language skills such as reading or writing English (Harris, 1969). This is because speaking is a very complex skill requiring the simultaneous use of a number of different abilities which often develop at different rates: pronunciation, grammar, vocabulary, fluency, the ease and speed of the flow of speech, and comprehension in oral communication. So, oral tests must provide a reasonably consistent assessment of the testee's performance to indicate his/her proficiency (Carroll, 1983).

The next question that needs to be addressed is what should be tested in an oral proficiency interview? The argument is basically between linguistic competence testing and pragmatic testing. In linguistic competence assessment a specific list or words and phrases is prepared for the examiner to listen for in the response to ensure dependable scoring (Lado, 1961). The raters all follow the same general procedures by devoting approximately the same time to the average interview: speaking to the candidate at about the same rate of speech, maintaining the same level of difficulty in the questions they ask and apply the same general rating standards (Harris, 1969).

Pragmatic assessments, on the other hand, do not specifically look for mastery of any particular grammar pattern or phrase, but measure communication effectiveness for real life situations (Oller, 1971). Students are not scored on their linguistic accuracy in elements such as pronunciation, vocabulary and grammar structures. Rather, they are scored on their ability to produce and comprehend messages. Errors in grammar and pronunciation are counted only if they interfere with the communication process (Vallette, 1977).

Arguments against discrete/linguistic type tests indicate that they are impractical for classroom teachers because of the (a) the scope of pre-testing needed on entry require elaborate test sheet questions which must be prepared and administered laboriously by already overburdened teachers;

(b) the statistical evaluation necessary to compute each student's level on pre and post tests can be troublesome; (c) the teacher who is most intimate with the program offered in the institute has no control over the subject matter being tested. So, discrete/linguistic type tests do not allow for spontaneous communication between the tester and the testee; rather, it is controlled communication (Carroll, 1982). Therefore, for placement of students in language programs, pragmatic testing seems to be favored because linguistic accuracy testing seems to limit the response of the testee which is unnatural, and there is little possibility for variation for both the tester and the testee.

Of the oral tests available at that time, perhaps the most respected by the experts was the Foreign Service Institute's Oral interview Test (FSI) (Jones, 1978). This is still true today. The FSI is a face-to-face conversation which is tape recorded and then later evaluated by two different raters using a scale ranging from 0 to 5. One of the main criticisms of this test comes from Wilds (1975) who says that the FSI is not ideal for normal academic situations "where all testing comes at once and where using two teachers to test each student would make it prohibitively expensive" (p. 34).

Another test, the Ilyin Oral Interview (1972), is used as a placement test for incoming ESL students. The interview consists of fifty items which are progressively more difficult. Pictures, too, are part of the test. These black and white pictures are a series of actions during a day in the past, today, and the future. Interviewees have to respond to questions. This test is simple, and has a low-increment scoring method which only requires one tester. However, the test is unable to handle low-proficient/level speakers.

Yet another test of importance to the design of the OPT is the Oral Production Test (Poczik, 1973). This oral test places students in one of three ESL levels and is based on the *Orientation in American English* series published by the Institute of Modern Languages. This would make the test appropriate if other texts were used. So, it was clear that even commercially available tests were not appropriate for the needs of the Institute. What was called for was a completely new test that would be sensitive to the context and the goals of the program.

Ten years on more these ready-made tests of oral proficiency, created by experts in the area of language testing, are available (for example, ETS' Test of Spoken English (TSE), the FSI is still in use, the Test in English for Educational Purposes, and many more). However, they suffer from the same problems that earlier oral proficiency tests experienced including "squishy prose descriptors" (Hieke, 1985, p. 140), poor reliability (Fulcher, 1987) and poor validity (Upshur & Turner, 1995). So the motivation for this researcher to make a new test that would suit the program's goals and the local context, that would be fairly reliable and valid in placing students was partly because of the inappropriate ready-made tests available at that time. It seems that the situation has not changed much in the past ten years for as Hughes (1989) and Weir (1990) point out teachers should still make their own tests because ready-made tests are inappropriate for certain contexts.

THE ORIGINAL STUDY

It was because of the above noted disadvantages of the discrete/linguistic type assessment procedures and the lack of appropriate ready-made tests of English oral proficiency that the author decided to redesign the test for the placement of students at the institute. Correct placement is crucial because research has indicated that when students are well placed at an appropriate level in the sequence of the program, the courses function more smoothly and the drop out rate declines (Vallette, 1977). Therefore, the main hypothesis of the study was that students whose English conversation skills are judged by means of an oral interview upon entry to an English program will be more likely to be placed at the correct level of difficulty for both understanding and speaking skills. Also, students that experience this oral interview will have a high percentage chance of passing to the next level of difficulty and will be less likely to drop out.

The original study employed the action research method of research (Best, 1977; Kemmis &

McTaggart, 1982) since it involved developing a new test and directly applying it to a real situation, namely, placing students at the Foreign Language Institute, Yonsei University, Seoul, Korea, in the fall term of 1985/86. The degree of success or failure of the new design would be demonstrated after it had been applied.

The sample group composed of 289 students of English as a Second Language (ESL) varying in age from 18 to 50 years old. All were native speakers of Korean. The interviews took place at the institute, and as a result of the interviews, the students were placed in a particular level reflective of their score. No written test was administered.

Ten instructors (native English speakers) were chosen to administer the interviews. Each instructor had at least two semesters teaching experience at the institute. Also, each instructor attended a training session in the methods and procedures of the new OPT. The training consisted of (a) listening to tapes of sample interviews and rating the tapes according to the new interview, (b) reviewing responses that would be acceptable for each particular level, and (c) presentation of the scoring techniques for the interviews and the relation between scores and levels.

Design of the Interview Instrument

The Oral Placement Test (OPT) was designed by the author to emphasize the communicative aspect of language rather than discrete point testing. This was deemed necessary in order to coincide with the goal of the Foreign Language Institute. The three tests discussed above, the FSI, the Ilyin Oral Interview and the Oral Production Test, had a profound influence on the design of the original instrument. The author included a special introductory section on grammar not to test grammar accuracy, but to set the interviewee at ease. Korean students have certain expectations as to what should be included in a test of English from their prior experiences and of all these, grammar testing is the main one. So, this interview would not disappoint them. The idea was to relax the interviewee at the beginning of the test. Then the test items were arranged at an increasing level of difficulty with the questions starting easy and getting more difficult in Section A, thus providing a 'warm-up' for the test proper.

Section B (which utilized a visual cue) and section C, was designed to emphasize communicative skills, with grammar patterns playing a secondary role and only having influence on scoring if the grammar mistakes impeded the interviewer's understanding of the context. In section B examinees were required to answer three questions: two questions directly related to the visual cue and the other related to the theme or topic of the picture. Scoring criteria included a description of the picture, a story about the picture, either fact or fiction, and an opinion, a reason for the opinion, and a justification for the reason. An overall score was given to the three responses.

Section C was designed to place emphases on creative, analytical and critical use of spoken English. Questions one to four are short questions with a clear topic to be addressed. Questions five through eight are longer and more complex topics requiring more accurate comprehension by the testees. Scoring criteria included an opinion, a reason for that opinion and a justification for that reason. For example, if all three criteria were met with minimum grammatical errors, four points would be awarded. However, if all three criteria were met many grammar errors that impeded the flow of conversation, then one point was deducted (see Appendix B for the scoring procedures).

An evaluation sheet had to be designed so that it was simple to score and would require minimal writing during the interview. Heaton (1975) says that there should be no marking of a testee during an interview as it can distract the testee. Harris (1969) says we should make a mental score during the interview and write it down after the testee leaves. So, the evaluation sheet consisted of scores form 0-5 for each section and the tester was required to circle one score in each section.

In order to test the success or failure of the OPT, a questionnaire was distributed to all the instructors at the institute. This questionnaire was designed to show the success or failure of the test

at more accurately assessing conversation skills of new students at the institute. Questions covered included the likelihood of new students advancing to the next level, the homogeneity of the student's spoken English ability and a separate evaluation of the former placement method (written test of English) as well as the new placement method (oral interview) at the institute.

Findings

Of the total sample group placed by the interviews, 75% passed onto the next level, 10% had to repeat the same level, and 14% left the institute. Also, through informal interviews with the author, all of the interviewees said they were very pleased with the new interview in that they had felt they had had a 'real' conversation with the testees and they were not really fatigued even after five hours of interviewing. They also indicated that they were pleased with the relative ease of the scoring procedures.

The results of the questionnaire administered to the students indicated that the instructors were comfortable with the placement of the students in their particular classes. They were also confident that most of their students would pass onto the next level because they had been placed at an appropriate level in the first place; that is with usual application to the study materials, they should progress to the next level. Regarding the question directly related to instructors' like or dislike of the OPT, only 10% expressed a dislike for the new interview system; the remaining 90% said it was excellent (30%) and good (60%). The results seem to support the initial hypothesis that well placed students have a greater likelihood of advancing to the next level and are less inclined to drop out.

THE OPT TEN YEARS ON

Michael Gibb, an instructor at the Foreign language Institute, Yonsei University, Seoul, carried out a critical assessment of the OPT in 1996 and found that "results, though encouraging, suggest a greater degree of placement accuracy is required" (Gibb, 1996, p. 1); he says that reliability and the criterion for levels of performance are potential sources of inaccuracy. Nevertheless, he found that the OPT does offer practical advantages.

Gibb outlined one of the major changes that the OPT has incorporated in the past ten years. Instead of the original one interview (in one go) with one interviewer, the structure of the OPT has been changed to two stages. Stage One involves an initial interviewer with what Gibb calls a 'sorter' which lasts one minute. The student is given a 'color-code' which indicates a rough estimation of the appropriate level. This color-coding system offers the main interviewer in Stage Two a chance to "fine-tune the initial estimation…Stage One is quite different than Stage Two since the 'sorters' don't have time to make an accurate decision—it's up to the interviewers to correctly place the student" (Gibb, 1996, p. 2). So, Gibb wanted to test the hypotheses that (a) "the OPT is failing to fulfil its purpose as shown by teacher dissatisfaction." (b) Hypothesis Two focused on the degree of agreement between the raters at Stage One and Two of the OPT:

> "If there are a significant number of disagreements between Stage One and Two, then there is a problem of scoring reliability. If there are a significant number of agreements, then there would seem to be a satisfactory degree of scorer reliability" (Gibb, 1996, p. 4)

Similar to the 1986 result, Gibb found that the teachers were relatively happy with the placement level of the students, but called for a greater degree of accuracy to place the students in the right class. He also found that Hypothesis Two showed "substantial agreement between Stage One and Stage Two of the OPT" (p. 6). When Gibb talked about a greater need for accuracy in the placement of students, he said: "This does not mean that the OPT is failing to fulfill its objectives, but that the process can be improved". He included the reliability of the scoring system and the criterion levels of

performance as the two areas that need improvement. For greater inter-rater reliability, Gibb proposes that two raters supply two test scores to be compared; "if the scores are consistent it implies that there is a degree of consistency. If the scores are significantly different, then the student can be retested, or referred to a third interviewer" (p. 7).

Regarding the criterion levels of performance, Gibb sees too much emphasis given to grammatical competency to place students into a communicative language program. He calls for new criterion categories that cover such aspects of communicative competence as accent, and appropriacy in order to test the candidate's overall communicative competence more accurately. However, as Canale and Swain (1980) point out not many can agree in the literature about whether or not the notion of communicative competence includes grammatical competence as one of its components. Munby (1978) however, contends that communicative competence should include the notion of grammatical competence because the opposite view could lead to two misleading conclusions: (a) grammatical competence and communicative need to be taught separately—first grammar, and (b) grammatical competence is not an essential component of communicative competence.

The two main advantages of the OPT, according to Gibb, are that it is practical and efficient. It is practical and efficient because of the use of a controlled interview and the use of two stages. The interview itself, he says, is uniform and consistent in question type, and the fact that 400 students take the test on any given occasion make the two-tier system of the interview more efficient. This is the main change from the original model devised by this author. In the original system, testees were only interviewed once, and their placement was based on that one test-score. Now Gibb says that this two-tier system is more reliable and valid.

FUTURE RESEARCH

One area for future research that may be an appropriate next step for the development of placement tests is a closer examination of oral tests themselves. For example, this researcher is aware that some programs place students by the use of a simulated interview in which a student, sitting in a language laboratory wearing earphones, speaks into a microphone and/or reads a passage and/or responds to some questions on a tape that were presented. The rationale for this method of 'oral' placement test is that it is easy to administer in terms of the few personnel required to carry out the test. However, this type of test, as opposed to face-to-face interview type oral placement test (like the OPT presented in this paper), may put some test takers at a disadvantage. As Bachman (1990) agrees, he says: "Some test takers, for example, may perform better in the context of an oral interview than they would sitting in a language laboratory speaking into a microphone in response to statements and questions presented through a pair of earphones" (p. 111). So, administrators of different language programs might consider the hidden cost of inappropriate placement of students with a 'cheaper' system. Research can be carried out with both types of tests administered to students and placement results compared.

Another, and more complicated, consideration is the gender of the tester; it is possible that in a culture like Korea certain test takers may feel more comfortable (which could lead to a better performance in the test) talking to a tester of a certain gender. This will, of course, vary from person to person and culture to culture. It nevertheless, may be an appropriate area for further research in a country like Korea. For example, it is this researchers experience that in classroom group discussions in Korean universities adult male Koreans were reluctant to talk 'openly' with adult female students and vice versa. However, this is an informal observation, hence the need for more formal research for the appropriate placement of English language students. There are other considerations too that could be researched in the Korean context such as the age of the tester, the personality of the tester—for example, will similar personality types (both test taker and tester) result in a better or worse performance?

CONCLUSION

Heaton (1975) says that speaking is far too complex a skill to permit any reliable analysis to be made for the purpose of objective testing. The main change in the OPT testing procedures of having two raters rather than one seems to have been effective in making the instrument more reliable. However, a certain amount of subjectivity is impossible to eliminate, since the variables of mood and emotion are human characteristics. Hughes (1989) says that accurate measurement of oral ability is not easy, but maybe two raters will be more reliable than one in most cases. Also, the original purpose of the OPT was this author's conviction that if a somewhat more formal oral examination was developed emphasizing a certain amount of communicative competence (Canale & Swain, 1980) and if it produced better results than written tests or grammar based interviews, then the test would be reliable enough for the purpose of placing students in appropriate levels. Also, as the purpose of the course is for students to learn how to speak English in conversational and business settings, the OPT presented in this paper seems to be an authentic (Bachman, 1990; Hoekje & Linnell, 1994) way to test the success potential of candidates by placing them at appropriate levels.

In 1986/7 the results indicated that the students were placed in correct levels, and that they had a high chance of progressing to the next level as a result of this proper placement. Also, the instructors were happy with the placement and the testers were not unduly fatigued with the interview process. Therefore, it seemed that the OPT fulfilled its role to be able to test/evaluate and then place a large number of people in as short a time as possible—five minutes or less. Even Valette (1967) agrees that this is a reasonable consideration: "If items are carefully chosen, the test itself can be quite short; much can be said in five minutes, much even in three" (p.121).

Now, ten plus years later, and with only one major change, the OPT still seems to be still accomplishing its goals. As Gibb (1996) says: "The results of [his] research project [that of critically analyzing the OPT] indicate that the OPT is fulfilling its purpose to a satisfactory degree" (p. 10). He sees the OPT as having two advantages: it is practical and it is efficient. Another point of agreement between these two studies is the success in placing students at similar levels of language proficiency; the original purpose of the OPT was to group students of similar ability levels together so teachers could focus in each class on learning (or problems) appropriate to students at similar levels of language proficiency. It also seems that ten years on the OPT still continues to do this also. This also supports the notion that EFL teachers can construct their own tests that meet their own needs and best suits their own context.

THE AUTHOR

Dr. Tom Farrell is a lecturer in the department of English and Applied Linguistics at the National Institute of Education, Nanyang Technological University, Singapore. Before that taught at the university level for sixteen years in Korea. He is editor of *The Korea TESOL Journal*.

REFERENCES

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Best, J. (1977). *Research in Education*. Englewood Cliffs, NJ: Prentice Hall

Brown, G. (1983). *Teaching the spoken language*. Cambridge: Cambridge University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Carroll. B. J. (1983). *Testing communicative performance*. London: Pergamon.

Clarke, J. (1978). Psychometric considerations in language testing. In B. Spolsky (Ed.) *Approaches to*

language testing. Arlington, VA: Center for Applied Linguistics.

Farrell, T. S. C. (1986). The Oral Placement Test. Unpublished Masters Thesis. University of Southern California.

Fulcher, G. (1987). Tests of oral performance: The need for data-based criteria. *ELT Journal, 41*(4), 287-291.

Gibb, M. (1996). A critical analysis of the OPT. Unpublished paper, University of Essex.

Harris. D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

Harrison, A. (1983). *A language testing handbook*. London: MacMillian Press.

Heaton, J. B. (1975). *Writing English language tests*. London: Longman

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Hoekje, B., & Linnell, K. (1994). 'Authenticity' in language testing: evaluating spoken language tests for international teaching assistants. *TESOL Quarterly, 28*(1), 103-126.

Ilyin Oral Interview (1972). Rowley, MA: Newbury House.

Jones, R. (1975). Testing language proficiency in the United States Government. In, R. Jones, & B. Spolsky (Eds.) *Testing language proficiency.* Arlington VA: Center for Applied Linguistics.

Kemmis, S., & McTaggart, R. (1982). *The action research planner*. Victoria, Australia: Deakin University Press.

Lado, R. (1961). *Language Testing*. London: Longman

Munby, J (1978). *Communicative syllabus design.* Cambridge: Cambridge University Press.

Oller, J. W. Jr. (1971). *Language tests at school*. London: Longman.

Poczik, R. (1973). *Oral production test*. Albany, New York: New York State Education Department.

Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal, 49*(1), 3-12.

Vallette, R. M. (1977). *Modern language tests*. New York: Harcourt, Brace, & Jovanovich.

Weir, C. (1990). *Communicative language testing*. New York: Prentice Hall.

Wilds, C. (1975). The oral interview test. In R. Jones & B. Spolsky (Eds.) *Testing language proficiency*. Arlington, VA: Center for Applied Linguistics.

<div align="center">

APPENDIX A

THE OPT INTERVIEW QUESTION SHEET (EXAMPLE)

</div>

Section A: Introduction (structure and grammar)

All questions in this section should be asked if the testees' grammar is suspect.

1. Please introduce yourself stating your name, age, education, family background and present occupation.
2. What do you like to do...in your free time?
3. What did you do last weekend?
4. What will you do after this interview?
5. How long have you been...married?
6. Report at least two things different people said to you last week.

Section B: Pictorial/ Opinion

Give the testee a picture.

1. Please describe the picture (what do you see?).
2. Make up a possible story for the picture.
3. What is your opinion about (topic related to the picture).

Section C: Creative, Analytical, and Critical use of Language.

Two questions should be asked from this section: one from 1-4 and the other from 4-8 because of increasing difficulty.

    1. If you went to live in another country, which of your customs would you like to keep and why?

    2. Do you think women should be required to serve in the military? Why or why not?

    ....

    5. What changes do you expect to occur in your country in the next ten years?

    6. Do you think parents should encourage a three or four-year-old child to begin studying something in order to try to speed up his or her mental development?

<center>

APPENDIX B

SCORING PROCEDURES FOR THE OPT

</center>

Students will be scored 0 to 4 points in each section, but please note two points: (a) pronunciation is not a factor unless it inhibits communication, and (b) listening comprehension should be considered in each section with one point being deducted if the testee continually asks for repetition of the question.

Section A: Accuracy

Correct grammar and correct information in the testees' answer are the main foci of this section. Accuracy means the correctness of the grammar in comparison to that of a speaker of standard English. Rather than consider the whole spectrum of grammar, attend to the structure, verb tenses, and subject-verb agreement in rough order of importance.

    4 points; Grammar is correctly used along with appropriate answers. Further, the testee volunteers more information than was required, and speaks without undue hesitation and with confidence.

    3 points; Grammar is appropriately used 90% of the time. The testee might correct errors. There could be some hesitation or a request for a question to be repeated.

    2 points; Grammar is 60%-90% accurate. There is more hesitation.

    1 point; There is a grammar error in almost every sentence.

    0 points; Every sentence has a grammar error, and the responses are very short. Or the testee has no idea how to answer.

Sections B and C: Communication

    The emphasis is on communication.

    4 points; The testee responds freely and without undue hesitation, continues when requested, and is highly productive; language flows until ideas are exhausted.

    3 points; The testee responds with one or more spontaneous remarks and continues with more when challenged by the interviewer. Makes some grammatical errors more like slips of the tongue.

    2 points; Similar response as 4 and 3 points however, makes many grammar mistakes, enough so that it could be distracting from comprehending the content of the answer.

    1 point; The testee does not respond until encouraged, and does not really continue after that. The tester must work too hard with the testee.

    0 points; The interviewer is just going through the motions

(Note: for ease of scoring, if the testee gets 1 point in section A and 1 point in section B then stop; otherwise continue)

Reviews

Second Language Acquisition.
Rod Ellis. Oxford: Oxford University Press, 1997.
Pp. x + 147.

Reviewed by Robert J. Dickey, Miryang National University.

Rod Ellis is one of the pre-eminent and most prolific writers in the multiple fields of Applied Linguistics, Second Language Acquisition, and Teaching English as a Second/Foreign Language. This release is the latest in a string of SLA texts/references he has produced in the past decade and a half. One might consider this an update and condensation of his previous works, which include the popular *Understanding Second Language Acquisition* (Ellis, 1985).

This paper-back book is designed as a well-documented introductory text to the field of Second Language Acquisition (SLA), a relatively recent off-shoot to studies in Linguistics, but now a recognized field in its own right. Part of the new "Oxford Introductions to Language Study" series (edited by H.G. Widdowson), this volume follows the series' format: separate sections in the book for surveying the field, brief readings in specific topics, annotated references, and a glossary.

Ellis is a favorite at TES/FL conferences for his direct, no-nonsense approach to communicating with his audience: this same approach is evident in this book.

On the other hand, *Second Language Acquisition* is not light reading. The publisher identifies it as a "brief survey" (back cover). I found that it can be read independently by newcomers to the field of SLA who are committed to independently learning the subject matter, or it could be used as a coursebook for an "Intro to SLA" course in a language teacher's program. Widdowson states in the Preface that "these surveys are written in the belief that there must be an alternative to a technical account on one hand and an idiot's guide on the other…" (p. ix).

It's not "an idiot's guide". The material demands due deliberation. New to the study of SLA myself, I couldn't really internalize more than one chapter in any given day. And it needed review, before proceeding to the next chapter. Those who are less familiar with current discussions in Teaching English as a Second/Foreign Language might well seek support of others, for example, forming a reading circle or discussion group, such as the KOTESOL Teacher Development & Education SIG did in November 1997 for Michael Wallace's (1991) *Training Foreign Language Teachers: A reflective approach.*

Well-summarized case studies are presented to clarify the discussion, and these same cases are utilized for multiple issues. Definitions within the text are complemented by more extensive definitions in the glossary, which also refer back to the point of first usage in the text. On the other hand, definitive statements of theory presented early in each chapter are sometimes over-encompassing. Modifiers of these statements are found only at the conclusion of each chapter, after each of the exceptions or competing theories have been discussed. This style of writing can be disconcerting to readers with differing learning strategies.

One example of this style of "modifying" is as follows: "We have seen that learner language is systematic" (p. 25). "These observations do not invalidate the claim that learner language is systematic, since it is possible that variability is also systematic" (pp. 25-26). "Research on variability *has sought to show* [italics added] that, although allowance should perhaps be made for some free variation, variability in learner language is systematic" (p. 30).

The Readings section is organized by Survey chapter and topics within each chapter. Each

"reading" is approximately 250 words, plus a brief introduction and full annotation, and is followed with questions for reflection. These readings are particularly helpful in understanding the various nuances of each issue through real- life models.

References are sorted by Survey chapter and topics within each chapter, and each has a comprehension difficulty rating: ■□□ is introductory, ■■□ is more advanced and technical, and ■■■ is specialized and very demanding.

This text largely focuses on the grammar-based area of Interlanguage. It could be argued that there are other, equally meritorious areas of interest in SLA which are largely left untapped. In six pages the differences of individual learners, which are described by Ellis as "psychological dimensions of difference", are briefly overviewed. This includes the important area of Learning Strategies. In the chapter Instruction and L2 Acquisition Ellis draws together these various issues, showing how teachers hope to benefit from the various fields of SLA, and that some SLA researchers have real concerns for applicability of SLA research in the classroom. The Survey's conclusion briefly discusses the problem with SLA encompassing such a wide variety of orientations and hypotheses, many of which may be mutually incompatible.

Teachers in Korea will enjoy using this book as a supportive first step in their inquiries of "how and why" students engage in certain tactics and share common approaches in learning, all without guidance from their instructors.

In short, this recent publication will be an invaluable addition to the reading list for practicing teachers who have not studied SLA, and would be a useful textbook for students engaged in an SLA-overview course.

THE REVIEWER:
Robert J. Dickey is a member of the Faculty of Liberal Arts at Miryang National University. His research interests include materials development for false beginners at the university level and "low-tech" resources for students and teachers.

REFERENCES
Ellis, R. (1985). *Understanding Second Language Acquisition* Oxford: Oxford University Press
Wallace, M. J. (1991). *Training Foreign Language Teachers: A reflective approach.* Cambridge: Cambridge University Press

Teacher Learning in Language Teaching
Donald Freeman and Jack Richards. (Eds.) Cambridge MA: Cambridge University Press, 1996. Pp. 386.

Reviewed by George Patterson, University of the Philippines

The book, *Teacher Learning in Language Teaching* edited by Donald Freeman and Jack C. Richards, contains sixteen chapters, which are divided into four sections. It introduces different perspectives on how language teachers and student language teachers respond to teaching and the experiences that are given as part of their professional development. In other words, it scrutinizes the process of learning to teach a second or foreign language, through descriptive accounts of the experiences of teachers. More specifically, this book illuminates the nature of learning to teach second or foreign languages through research-based accounts of how teacher education programs and the experience of teaching mold the knowledge, thinking, and the practice of language teachers.

The editors do not restrict "second language" to English since they believe that learning to teach any language shares certain basic characteristics.

This book makes a fundamental point that in order to better comprehend language teaching, one needs to know more about language teachers: "what they do, how they think, what they know, and how they learn." Specifically, they need to comprehend " more about how language teachers conceive of what they do: what they know about language teaching, how they think about their classroom, and how that knowledge and those thinking processes are learned through formal teacher education and informal experience on the job."

The chapters are arranged in three sections: In Section I, five chapters explain the start of teacher learning. In the first chapter, Kathleen Bailey and her colleagues explain their employment of language learner autobiographies as a medium for examining their own professional development. They demonstrate how writing about past experiences of teaching and learning can serve as a powerful stimulus for additional learning, disclosing implicit assumptions and beliefs about the nature of teaching. In the second chapter, Karen Johnson investigates the experience of a student teacher during a teaching practicum, and demonstrates how this experience molds the teacher's understanding of herself as a teacher, of second language teaching, and of the practicum itself. The gap between her vision of teaching and the practical realities of the classroom form a tension which interferes with her perception of what her students are learning. In Chapter 3, Gloria Guttierrez Almarza concentrates on the relationship between student teachers' background knowledge and the knowledge that they receive in teacher education courses, and examines how these sources of knowledge interact during teaching practice. Anne Knezeviv and Mary Scholl explain their experience of collaborative teaching in a graduate Spanish course, and demonstrate how collaborative learning, teaching, and reflection molded their understanding of themselves and teaching.

In Section II, there is a focus on the practice of teaching itself, and on the cognitive process that teachers participate in as they develop expertise in teaching. Patrick Moran scrutinizes the role that models of teaching can play in a self-directed process of learning to teach, and recommends that they serve as temporary learning strategies as teachers from their own personal teaching paradigms. Anne Burns scrutinizes six experienced Australian teachers who were faced with a new teaching situation—teaching adult learners. She scrutinizes how their pre-existing beliefs influenced their classroom practices and how subsequent changes in their beliefs happened. She explains the interactions among institutionally derived beliefs, personal beliefs and thinking, and the process of instruction itself, and demonstrates how top-down and bottom-up processes interact in a process that she alludes to as "intercontextuality". The third chapter in Section II, by Polly Ulichny, scrutinizes a segment of a lesson to clarify the contribution of context and personal interpretation in the teacher's resolution of a teaching. Her chapter demonstrates how analysis of the discourse of teaching can give insight into teacher's cognitive processes. In the final chapter in this section, Deborah Binnie Smith scrutinizes a group of secondary ESL teachers in Canada, the instructional decisions that they make, and the factors that affect those decisions, demonstrating that teacher decisions are not isolated or arbitrary but part of a complex, interrelated process which is informed by beliefs, perceptions, experience, and context.

The chapters in the third section of this book explore the relationships between teacher education and teacher learning. In Chapter 10, Donald Freeman explains how a group of teachers in an in-service master's program incorporate new ideas into their thinking about classroom practices. His analysis traces the ways in which the teachers rebuild their classroom practice, employing professional discourse to rename their experiences, and therefore assign new or different meanings to their actions. In Chapter 11, Jack Richards, Belinda Ho, and Karen Giblin depict a group of teachers in an initial training course, and demonstrate how their individual conceptions of teaching result in different concerns within lessons as well as to different perceptions of what makes those lessons more or less successful. They also show that what trainees learn from a program is not simply a

reflection of the program's component. In Chapter 12, Francis Bailey explains a collaborative approach to the methods class in a graduate teacher education program. He analyzes discourse sequences from interactions within small group discussions, which disclose how the instructor capitalizes on the heterogeneity of the group to achieve shared dialog among the participants. In the following chapter, Michael Wallace scrutinizes the role of the professional project within teacher education and, by following the use of such projects in a B.Ed. program, raises questions on the nature of action research and its suitability within certain kinds of teacher education programs.

In Section IV in the final chapter of the book, Donald Freeman reviews the field of teacher cognition and teacher learning within which the research in learning to teach second languages is arising. He puts the work in the context of general educational research, tracing its antecedents both in current educational research and in research in second language teaching. He then investigates the central conceptual issues confronting such work and the types of research methodologies that are illustrated in the studies mentioned in this book. The chapter gives both a review of the issues investigated in the individual chapters as well as a framework for this new area of research.

This book is an excellent book for language teacher education development of second language teachers, EFL/ESL teachers, and Second Language Teachers in terms of theory and practice, in-service and pre-service training. It also shows how second language teachers integrate and employ the theoretical ideas and theories as well as the pedagogical principles that they acquire during professional education.

The Korea TESOL Journal

The Korea TESOL Journal welcomes practical and theoretical articles concerned with any aspect of foreign language learning and teaching in international contexts, especially Korea and Asia. Areas of interest include

1. classroom-centered research
2. second language acquisition
3. teacher training
4. cross-cultural studies
5. teaching and curriculum methods
6. testing and evaluation

The Korea TESOL Journal prefers that all submissions be written so that their content is accessible to a broad readership, including those individuals who may not have familiarity with the subject matter addressed. *The Korea TESOL Journal* invites submissions of full-length articles (4 copies) of no more than 20 pages, including reference list, typed on A4 paper, and double-spaced. To facilitate the blind review process, authors' names should appear only on a cover sheet, not on the title page; do not use running heads. Manuscripts should be submitted to the Editor of *The Korea TESOL Journal* (see address below). *The Korea TESOL Journal* also invites short, evaluative reviews of professional books, classroom texts, computer software, and tests. Reviews should provide a descriptive and evaluative summary and a brief significance of the work. Submissions should be no longer than 500 words. Submit two copies of the review to the Review Editor (see address below).

Style

The Korea TESOL Journal follows the style of the *Publication Manual of the American Psychological Association (4th ed.)* (APA). Please see a recent copy of *TESOL Quarterly* for examples of documentation and reference lists. Copies of APA style guidelines may be ordered from APA Order Department P.O. Box 2710 Hyattsville, MD 20784 United States. This is a strict requirement. Please put a word count at the end of the manuscript. Please submit the following:

1. Abstract (no more than 200 words)
2. Four copies of the manuscript
3. Biographical sketch(es) (no more than 50 words)

Evaluation procedure is usually completed within two months.

Restrictions

Papers sent to *The Korea TESOL Journal* should not have been previously published, nor should they be under consideration for publication elsewhere. We regret that manuscripts cannot be returned.

Submissions

Send submissions to:
Korea TESOL Journal
P.O. Box 391 Seo-Taejon, South Korea 301-600

Membership Application / Change of Address

Please fill in each item separately. Do not use such timesaving conventions as "see above." The database programs used to generate mailing labels and membership directories sort answers in ways that make "see above" meaningless. Long answers may be truncated. Use abbreviations if necessary.

Please check the items that apply to you   Type of Membership

☐ Individual (40,000 won/year)

☐ New membership application   ☐ Educ. Institutional (80,000 won/year)
☐ Membership renewal   ☐ Commercial (600,000 won/ year)
☐ Change of address / information notice   ☐ International (US$50.00 / year)
☐ Life Time (400,000 won/ year)

Payment by   ☐ Cash   ☐ Check   ☐ On-line transfer (Please make on-line payments to Yeom Ji-sook or KOTESOL at Seoul Bank account number 24701-1350101. In order to insure that your application is properly processed either include your name on the bank transfer slip or send a photocopy of the bank transfer slip with this application.

First name: _____   Last name: _____
Chapter: _____ (Seoul, Taejon, Taegu, Cholla, Pusan, Cheju, Chongju, International)
Date of membership: _____-_____-_____ (Today's date.)
Confidential: _____ (YES or NO) *(If you answer YES, the following information will not be included in any published form of the membership database. The information will be used by KOTESOL general office staff only for official KOTESOL mailings.)*

E-mail address: _____
Home phone (_____)_____ fax (_____)_____ beeper (_____)_____
Address:_____
_____
_____ Zip: _____-_____
Work phone (_____)_____ fax (_____)_____ beeper (_____)_____
Address:_____
_____
_____
_____ Zip: _____-_____
To which address would you prefer KOTESOL mailings be sent? _____ (Home / Work)

Please check those areas of ELT that interest you:

☐ Global Issues   ☐ CALL   ☐ Teacher Development
☐ Reading/Writing   ☐ Elementary Education   ☐ Post-secondary Education
☐ Intensive English Programs   ☐ Adult Education   ☐ Learning Disabilties
☐ Teaching English to the Deaf   ☐ Speech/Pronunciation   ☐ Inter-cultural Communication
☐ Applied Linguistics   ☐ Video   ☐ _____

Date_____ Signature:_____

Mail or fax this form to:
Tony Joo, General Manager c/o KOTESOL Central Office P. O. Box 391-SeoTaejon 301-600
(Tel) 042-255-1095 (Fax) 042-255-1096 (e-mail) <kotesol@chollian.net>.

98-journ