# Language Policy Implementation: A Look at Teachers' Perceptions

**Hong Wang**
*Faculty of Education, Mount Saint Vincent University, Canada*

**Bio-Data:**
Hong Wang (Ph.D.) is an assistant professor in teaching English as a second/foreign language at the Faculty of Education, Mount Saint Vincent University, Canada. Her research interests are ESL/EFL teacher education and professional development, teaching English as a second or foreign language, language policies, and curriculum implementation and evaluation. Some of her studies have been published in Teaching and Teacher Education, Asian EFL Journal, and TESL-EJ. Before joining Mount Saint Vincent University in 2006, she was a SSHRC (Social Sciences and Humanities Research Council of Canada) Postdoctoral Fellow in the School of Linguistics and Applied Language Studies at Carleton University, Canada.

## Abstract
This study explores teachers' perceptions of the language policy implementation in the Chinese tertiary context. With data collected from classroom observations and follow-up interviews, the findings revealed a discrepancy between policymakers' intentions and teachers' executions. Teachers failed to implement faithfully what was required from policymakers in the classroom. Rather, they conducted teaching based on the classroom and political realities. Their factors were mainly student and departmental, which included large class size, students' language proficiency, motivation, learning behaviour, perceptions of teachers' role, and institutional evaluation mechanism on teaching excellence. The implications of this policy implementation point to the importance of understanding why there exists such a discrepancy within that context.

**Key words:** teachers' perceptions, language policies, implementation, discrepancy, classroom reality, college English teaching.

## Introduction

Curriculum policy implementation in any educational jurisdiction involves a variety of stakeholders. Their roles in executing the curriculum policies contribute to the degree which new or revised curricula will be implemented in the local institutions. Researchers have been cognisant that teachers as implementers are the most important players. Studies have also demonstrated that implementers did not always do as told nor did they always act to maximize policy objectives (Cohen & Ball, 1990; McLaughlin, 1987). Moreover, teachers have often been

diagnosed as "resistant to change," or just simply lazy when they ignored or subverted curricular innovations (McLaughlin, 1987; Smit, 2005). Spillane, Reiser, and Reimer (2002) looked at their situation in a different light, explaining that this is because implementers often lack the capacity - the knowledge, skills, personnel, and other resources - necessary to work in ways that are consistent with policy. Spillane et al. warned that even if implementers construct understandings that reflect policymakers' intent, they may not have the necessary skills and resources to do what they understand the policy to be asking of them.

Wang and Cheng (2005) concurred, saying that teachers' failure to implement policy as policymakers hoped may signal their uncertainty about outcomes and their assessment that new practices are not as good as the previous ones. Gross et al. (1971) found that teachers' will to implement the imposed change declined over time because the change was not supported by their education director in ways that impacted them. Through their empirical study on English language instruction in classrooms in China, Japan, Singapore, Switzerland, and USA, Silver and Skuja-Steele (2005) examined how policy and classroom practice interact by comparing classroom practices and teachers' statements of pedagogical rationales with governmental policies. They found that teachers were aware of policy initiatives related to language education. However, teachers were focusing on immediate classroom priorities that influenced daily lessons and put their emphasis on student learning. Their findings revealed that language policies were reinterpreted into structural priorities which indirectly influenced classroom priorities and so filtered through to classroom practice. Furthermore, teachers' willingness to implement language policies was influenced by the social and personal dimensions of classroom teaching and by teachers' goals and beliefs. In other words, changes were mitigated by the local (contextual) factors.

Researchers (Gross et al., 1971; Spillane et al., 2002) have discussed impediments to implementation and reasons of why implementation fails in actual practice on the part of teachers, the implementers. In summary, these obstacles are: teachers' lack of clarity about the innovation; lack of knowledge and skills needed to conform to the innovative initiative; unavailability of required instructional materials; incompatibility of organizational arrangements with the innovation; lack of staff motivation; teachers' prior knowledge, beliefs, and experiences; different interpretations of the same policies; and misunderstanding or superficial understanding

of the policies. All of these reasons potentially impede teachers from being able to implement the intended curriculum policies.

**The Research Context**

This study[1] explored two teachers' perceptions of the national English as a foreign language (EFL) curriculum policies. It was conducted in the Chinese tertiary setting, where two types of English language education exist: one for English major students and the other for non-English major students (Wang, 2001). Language education for English majors centres on developing students' language proficiency to advanced/sophisticated level; only a small number of university students are enrolled in such a program. Language education for non-English majors is called "College English," which refers to the English language instruction in both universities and colleges. Non-English majors constitute the largest proportion of tertiary-level students pursuing undergraduate degrees in a variety of disciplines such as arts, sciences, engineering, management, law, and medical science. These students study English primarily as a tool to help them achieve advancement in their own fields. In 2003, approximately 3.1 million students (http://www.china.org.cn) were enrolled in such English instruction at 1,571 Chinese universities and colleges (http://www.edu.cn) in 32 provinces, autonomous regions, and municipalities.

Under the guidance of College English Teaching Syllabus (College English Syllabus Revision Team, 1986, 1999)[2], all college English students are required to study English for two years. They take a total of 280 teaching hours of English (about 70 hours each term or 5 to 6 hours each week) in order to meet the basic requirements. In terms of textbooks which represent the syllabus, dictate what needs to be taught, and "are compiled by a government-appointed panel of experts" (Wang, 1999), college English teachers are expected to closely follow the structure of textbooks to plan their teaching. Moreover, to examine the implementation of the curriculum, students are assessed using a nationwide, standardized English proficiency test called the College English Test Band 4 (CET-4) after the first two years of English study. The CET-4 focuses on testing students' language proficiency in listening, reading, and writing. Most of the test items are in a multiple-choice format.

**Methodology**

*Sampling*

Using purposive sampling (Patton, 2002), I chose a veteran senior teacher and a novice young teacher for the classroom observations. The intention was to ensure that the selected two teachers could provide "information-rich" cases with respect to the purpose of the study. I expected them to have a sound knowledge and better understanding of college English teaching in China. The sampling started upon my arrival in Xi'an, the research site. I first contacted the department heads via telephone and invited them to recommend teachers who might be interested in my study and requested their contact information. In lieu of ethical consideration (of a power distance relationship between the administrators and their teachers), I asked the department heads or the directors of the teaching and research groups to provide a pool of potential subjects. Then from that list, I made phone calls or approached teachers directly after I delivered my research seminar in their university and eventually decided upon two teachers as my participants for this study.

*Instrumentation: Observations and Interviews*

*Research question*

Classroom observations and follow-up interviews in this study were intended to address the following research question: How is the intended curriculum interpreted by classroom teachers? From classroom observations, I saw how much teachers adhered to the objectives of the 1999 College English Teaching Syllabus (the 1999 Syllabus thereafter) in their classroom teaching activity; how teachers conducted their teaching; how much the target language, English, was used; and whether the nationwide College English Test exerted any impact on teachers' classroom teaching. From interviews, I wanted to find out why teachers conducted teaching in the ways that they did. In so doing, I aimed to examine how teachers actually conducted their teaching to achieve the intended curriculum.

*Observation guide*

I designed an observation guide (see Appendix A) to facilitate my classroom observations. This guide referred to Fröhlich, Spada, and Allen's (1985) COLT (Communicative Orientation of Language Teaching) category, comprised of three main features: (1) participant organization, (2)

4

activity type, and (3) use of target language. Participant organization describes who is conducting most of the talking in terms of teacher talk and student talk. Activity type focuses on what kind of teaching activities are conducted in the classroom to promote students' language learning. Use of target language identifies how much mother tongue and target language are used respectively in classroom teaching and learning.

I referred to the COLT observation model (Fröhlich et al., 1985) because it embraces more communicative orientation on what teachers and students do in the classroom and how they interact with each other. Drawing upon Part A of COLT, I described classroom events at the level of activity. This was particularly essential in my investigation of the college English curriculum implementation because the 1999 Syllabus laid special emphasis on communicative activities. Drawing upon Part B, I focused on the use of target language (L2) and the use of mother tongue (L1).

### *Interview protocol*

The instrument used in interviews with teachers followed what Patton (2002) referred to as the interview guide approach. In this approach, I listed the questions to be explored in an interview and used the list as a guide to "ensure that the same basic lines of inquiry are pursued with each person interviewed" (p. 343). I did not have to follow these questions one by one during the interview in any chronological order (Johnson & Christensen, 2004). Rather, this interview guide provided the topic dimensions associated with syllabi, textbooks, and tests within which I was "free to explore, probe, and ask questions that will elucidate and illuminate that particular subject" (Patton, 2002, p. 343).

The interview protocol (see Appendix B) with the two teachers was divided into eight dimensions. They were: (1) demographic information about teachers' language learning experience, educational background, and teaching experience; (2) clarity and appropriateness of the syllabus; (3) the college English teaching syllabus; (4) college English textbooks; (5) the college English test; (6) teacher training support received from the department; (7) rationale of English and Chinese use in teaching and the student-centred approach, and (8) challenges faced.

*Data Collection*

I adopted what Patton (2002) referred to as "unobtrusive observations" (p. 291) or non-participant classroom observations in collecting observation data. Out of ethical considerations, I first informed the two teachers of my research purpose and obtained their permission, which allowed me to enter their classroom to collect observation data. Then I worked out an observational schedule with each individual teacher. Final arrangements were confirmed about the location of each class period and the content of their instruction.

Classroom observations with the two teachers lasted from June 7 to 23, 2004. Together, five classroom observations were carried out with each teacher, Lily and Sally (pseudonyms), within a span of two weeks. During each time slot, there were two class teaching periods, each lasting 50 minutes. I observed one unit of the "Reading and Writing" (eight class teaching periods) and one unit of the "Listening and Speaking" component (two class teaching periods). To facilitate observations and remedy what is lacking through audio recording, I took field notes to record what I heard and observed as each lesson progressed. I recorded my reflections following each observation session to minimize any disruptive influence that I might have on the classroom environment.

Two face-to-face follow-up interviews were also conducted with each teacher. The initial one was an informal 15-minite conversational interview, which asked demographic information about their education, learning experience, and teaching experience. The second interview was a formal one, more structured and in-depth. It centred on teachers' experience in following the 1999 Syllabus and their perceptions of various issues with regards to the curriculum implementation in their classroom. Particularly, this interview focused on clarifying questions or inquiring about teachers' rationale for employing various strategies in their teaching. Each formal interview lasted 50 to 60 minutes and was conducted mainly in Chinese (Mandarin) only with English words, phrases, or sentences added for the purpose of clarification and elaboration or when the interviewees felt a need. The interviews were held in the staff meeting room on their university campus. The two EFL teachers were given pseudonyms, coupled with an indication of whether they taught English at a large or small university. With their permission, the interviews were tape-recorded.

*Data Analysis*

Analysis of the observation data in this study involved combining readings of my field notes with listening to audio-taped sample lessons. Following the observation guide, I conducted the analysis as follows. In the first stage, I listened to all the lessons recorded from the two observed EFL teachers and examined my field notes. In the second stage, I reviewed my detailed field notes coupled with interviews and identified themes that emerged. I described and supported these themes with evidence from both field notes and interview transcripts. In the third stage, through analysis of lessons recorded and field notes, I demonstrated how the curriculum policies were interpreted by the two teachers.

Interview data analysis adopted the "interpretive model" (Hatch 2002, p. 179) in the qualitative paradigms. By interpretive, Hatch (2002) referred to "a way to transform data that emphasizes interpretation" (p. 180). According to Hatch and other scholars as well (Denzin, 1994; Patton, 2002), interpretation is a defining element that permeates all qualitative research through making inferences, developing insights, attaching importance, refining understandings, drawing conclusions, and extrapolating lessons. Since researchers carry out interpretations in the research process, they make sense of the phenomenon under investigation. This interpretive analysis was intended to link interpretation to the data in order to result in meaningful data. Therefore, I followed what Hatch (2002) delineated as the steps in the analysis of the interview data. First, I read the data, immersed in them to get a sense of overall impressions. Then I reviewed the impressions and identified and recorded them in memos. After studying the memos for salient interpretations, I read the data again, coding places where interpretations were supported or challenged. Finally, I identified the excerpts that supported the interpretations.

**Findings**

The findings resulted in a number of themes which I have summarized and grouped into the following four emergent themes: adherence to and knowledge/understanding of the syllabus, the learner-centred approach, use of the target language, and impact of the CET on teaching. Below is the presentation of each of these themes.

*Adherence to and Knowledge/Understanding of the Syllabus*

It was set out in the 1999 Syllabus, that "College English aims to develop in students a relatively high level of competence in reading, and an intermediate level of competence in listening, speaking, writing, and translating so that students can communicate in English" (College English Syllabus Revision Team, 1999, p. 1). With the syllabus objectives as a reference, both Lily's and Sally's teachings are reported through activity types in the classroom. According to Fröhlich et al. (1985), the activity type describes what kind of teaching activities are conducted in the classroom to promote students' language learning. Through these activities, I demonstrated whether or not the two teachers followed the syllabus to cultivate students' five language skills.

On the whole, the classroom observations revealed that both Lily and Sally's teaching focused more on developing students' receptive skills of reading and listening than on promoting students' productive skills of speaking and writing. These two teachers conducted classroom teaching exclusively in accordance with the textbooks assigned to them instead of adherence to the objectives set out in the syllabus. Cultivation of students' reading skills was fully covered in their teaching activities. This was reflected in both teachers' detailed explanation of texts and of reading skills.

In terms of listening skills, both Lily's and Sally's students received sufficient practice. Translation practice was common in both Lily and Sally's teaching activities. While Sally involved her students in English to Chinese translation exercises, Lily conducted both English to Chinese and Chinese to English to engage her students in pattern drill practice and to check whether students understood the text or not.

However, compared with the time spent on cultivation of students' reading and listening skills, writing and speaking skills were underrepresented in Lily and Sally's teaching, particularly speaking skills. Lily conducted a few activities in which she encouraged her students to practice oral English. For example, Lily organized one speaking activity "asking about people's opinions and giving opinions" in "Listening and Speaking" class. She asked students to listen to a dialogue first and pay attention to the communicative function. Then she assigned students in pairs working on "opinion giving" about two social problems. Three pairs of students were asked to present their pair-work to the whole class. However, during my observations of her ten teaching periods, Sally did not organize any speaking activity at all except that she engaged her students in choral repetition of sentences and of listening passages in both "Reading and Writing" and "Listening and Speaking" components.

The over-representation of receptive skills of reading and listening in Lily and Sally's instruction seemed to be in line with the requirements of the syllabus, thus being faithful to the intended curriculum proposed by policymakers. However, the under-representation of productive skills of writing and speaking prevented them from completely implementing the proposed curriculum. In my opinion, speaking in particular was least emphasized in classroom activities in both Lily's and Sally's teaching.

Although the national policymakers expected teachers to adhere to the objectives and the specifications of the syllabus in their teaching and to be knowledgeable and clear about the syllabus, interviews with Lily and Sally revealed that these two teachers failed to meet the expectations. They claimed that they did not have a sound working knowledge of the 1999 Syllabus, nor did they have a deep understanding of it, although each of them was given a copy by their department head. Both teachers said that they only had very "vague" idea about the language skill requirements in the syllabus. Lily told me that her whole impression of the 1999 Syllabus was that it focused on reading and emphasized on input (i.e., reading and listening) rather than output (i.e., writing and speaking).

Sally expressed her lack of interest in this curricular document. She said, "Although there are several versions of syllabus, I'm not interested in it and I don't read much either. … every time I was assigned to teach new textbooks, I would teach in the same way as before" (TEIN01SU04)[3]. For her, there was no change in her teaching because of syllabus change. Similarly, Lily expressed her unawareness of the syllabus:

> I really don't understand much of that syllabus. So if you ask me, I really feel, I don't know much of the syllabus, so I don't have much to say. Many years ago, we were given this syllabus. I remember we had teachers talking about it. But to tell you the truth, I never read it. … I should say, I don't think I have a good understanding of it, because seldom did I study it or read it. It's a waste of my time. … I don't take it seriously, and maybe, ok, for personal reasons, I don't like such kind of thing. I want to get in touch with something really concrete and specific, that kind of thing. (TEIN02LU01)

### The Student-centred Approach

The national policymaker in terms of textbook implementation clearly stated that language teaching and learning in the classroom should centre on students, should reduce teachers' speaking time, and should encourage student participation. The observations from Lily and Sally's classes revealed that both teachers had adopted a teacher-centred approach. To see how

their classroom teaching was conducted, I reported on both Lily's and Sally's lessons through participant organization, a parameter to describe basic patterns of classroom interactions between teachers and students (Fröhlich et al., 1985, p. 53). The rationale was to find out who was speaking the most in the classroom, particularly by comparing teacher talk (indicating teacher-centred teaching) versus student talk (indicating a student-centred approach). I mainly observed two patterns: (1) Teacher work: Is the teacher lecturing to the whole class? and (2) Student work: Are students working chorally, in groups, or individually? (Fröhlich et al., 1985).

In the "Reading and Writing" component, the teachers' speaking was mainly lecturing to the whole class. Teacher talk took up 70% of the class time and student talk 30% in Lily's class. Likewise, teacher talk took up 85% and student talk 15% in Sally's class. Students' work in both teachers' classes covered choral work such as answering teachers' questions or checking multiple-choice exercises together, oral presentation, and individual seat work such as reading vocabulary and paragraphs aloud.

In the "Listening and Speaking" component, both Lily and Sally spoke less than they did during their "Reading and Writing" component. Both of them spent most of their class time engaging their students in listening practice, with Lily taking up 80% of the time and Sally 85%. Teacher talk only took up 8% in Lily's class time and 3% in Sally's when they explained new words in the listening passages in either English or Chinese. Although this course was designated as "Listening and Speaking," students' speaking seemed scant. Lily conducted a speaking activity of "giving opinions" and asked students to give oral presentations to the whole class. Her students' pair work preparation and oral presentations took up 10% of the class time. Sally, on the other hand, did not organize speaking activities in the class. Instead, she asked the whole class to repeat the listening passages chorally. She explained that the limited class time was so precious that only choral work (12%) would enable students to have more chances of practicing oral English.

Although the textbooks that Lily and Sally used advocated the "student-centred" approach, observations of their instruction revealed the lack of teachers' fidelity to this practice. When asked why the expected model from the policymakers was not implemented in their classrooms, both Lily and Sally expressed helplessness and listed reasons for not being able to do so. The two teachers listed large class sizes (over 55 students in each class), students' low language proficiency, insufficient teaching periods (five hours per week), heavy workloads each term

(completing required teaching tasks), and Chinese students' study habits (depending too much on teachers for instruction) as obstacles.

The above results demonstrated both teachers' non-implementation of the intended curriculum in their actual classroom teaching. Instead of following what was expected by policymakers, namely the learner-centred approach, both teachers primarily adopted a teacher-centred approach with teacher talk taking up most of their class time. The interviews also indicated that it was not that these teachers did not want to change their focus to student-centred learning, but that the teaching reality constrained them from carrying out activities centring on learners. Teachers' implementation effort was in fact untenable in practice.

### *The Use of English in Classroom Teaching*

The policymakers expected teachers to use English—the target language—as much as possible in classroom teaching. The rationale was the advocacy in the syllabus of creating a favourable language environment so that students would be immersed in an English-speaking context (College English Syllabus Revision Team, 1986, 1999). One policymaker even suggested using English entirely and using Chinese very little or not at all in teaching.

The classroom observations revealed that both Lily and Sally used the target language of English (i.e., foreign language or L2) and mother tongue of Chinese (i.e., first language or L1) in every teaching period of 50 minutes in "Reading and Writing" as well as in "Listening and Speaking." Both teachers used over 80% of English when explaining the text in "Reading and Writing" and used 90% of English in "Listening and Speaking." Both teachers used Chinese when checking multiple answers in students' workbook, explaining students' assignments, and translating some difficult sentences.

The observations revealed that both Lily and Sally had a lot in common in the use of L1 and L2. In the "Reading and Writing" class, both Lily and Sally used English (L2) when lecturing on the whole text, explaining the text structure, describing the organization of one unit, paraphrasing difficult sentences, and asking students comprehension questions. Both of them used Chinese (L1) when translating English sentences, analyzing grammatical sentences, and checking answers to multiple-choice exercises. Likewise, their students used English when answering teachers' questions and checking answers to multiple-choice exercises. Their students only used Chinese when they were asked to do translation exercises. In the "Listening and Speaking" class,

both Lily and Sally used English to explain new words and conduct classroom teaching. They used Chinese when they felt a need to explain some words or sentences. Their students only used English when checking answers of true or false or multiple-choice exercises in addition to passage listening.

Although English was encouraged by policymakers to be maximally exposed to language learners, the observations revealed that Lily and Sally still used a large amount of Chinese in teaching EFL learners. In the interviews, both of them indicated that language teachers should use more English in their teaching. However, they admitted that they had to use a lot of Chinese. They listed the following reasons for their choice: saving time; for clarity in text explanation; and checking students' understanding of the text.

### *The Impact of Tests on Curriculum Implementation*

Although Lily and Sally were instructing freshmen who would attend the nationwide CET-4 the next year, the observations revealed that testing impact was still apparent in both teachers' classroom instruction. The textbooks that Lily and Sally used in "Reading and Writing" and "Listening and Speaking," particularly students' practice books, contained a large proportion of multiple-choice exercises. Therefore, both Lily and Sally spent considerable classroom time checking comprehension questions about the reading, and doing grammar and vocabulary exercises and cloze tests.

The intended curriculum by the policymakers emphasized that the CET was voluntary for universities and students, that college English teachers should follow the syllabus, and that testing should bring forth positive effects on classroom teaching. However, in the interview, both Lily and Sally expressed strongly that classroom reality forced them not to implement what was intended by the policymakers.

Because the CET-4 was mandatory for Lily and Sally's universities and their students, these two teachers concurred that the most powerful influence of the CET on their implementation endeavour was their teaching to the test. Rather than just following the syllabus, both Lily and Sally taught what was tested in the CET. Sally claimed, "There is no impact on my teaching whether there is a syllabus or not. My teaching is primarily influenced by the CET, this 'magic rod.' So testing is our magic rod" (TEIN01SU06). She further admitted,

> What we are teaching is following the "magic rod" of the CET. It is definitely true. No matter what is required in the syllabus, and if there is no requirement in testing, nobody pays attention to it [the syllabus]. … If there is a gap between the syllabus and the test, we definitely follow the test. (TEIN01SU04)

They illustrated how they taught what was tested in the CET. One salient example, Lily cited, was that she paid special attention to students' writing, because the CET-4 set the minimum requirement in writing. This meant that students would not pass the test if they failed to get six points out of 15 points on the writing part of CET-4. As a result, she assigned students paragraph or passage writing as homework, which was consonant to the writing format in the CET-4. Another example was that after the original multiple-choice questions in listening comprehension were changed into dictation, Sally modified her teaching accordingly by adding dictation practice in helping her students to spell English words correctly. During the classroom teaching I observed, both Lily and Sally kept reminding their students what was tested in the CET. They would tell their students which language points were tested in a certain year. Sally explained,

> Unconsciously, I am doing this all the time, which means that I'm attaching really great importance to this test. I'm telling you the truth. I've never told my students that our syllabus requires you to do this or that. I've never said that. I would tell my students, "name after," this verbal phrase was tested in a certain year. I also tell my students that "available" is an important word because it was tested twice in the CET-4. (TEIN01SU05)

The interviews with Lily and Sally also revealed that the high stakes associated with the CET test undermined their effort in cultivating students' communicative abilities, for the reason that speaking skill was not tested for most students, although advocated in the intended curriculum. These two teachers saw this inadequacy as "problematic" and "the worst thing of the test impact," because "even when students can get high scores and good grades in the examinations, seldom do they write fluently or speak fluently" (TEIN02LU04). They criticized the short-sighted practice of teachers' disregarding students' communication skill but felt vulnerable at the same time. Lily explained, "while there is some conflict, I should say, I will place more emphasis on the language points which are associated with the examination [i.e., the CET]. It has something to do with the test for sure" (TEIN02LU03).

Lily conveyed that the CET exerted an indirect impact on her teaching through students' evaluation. She received considerable complaints from her students because she organized too many speaking activities in class! Her belief of language teaching was "to create a classroom in which my students can practice the English language, or communicate or discuss in the language" (TEIN02LU05). However, this belief seemed impossible for her to realize. She said that each term students evaluated their teachers based on how teachers helped them pass the CET. Her department and university attached greater importance to such evaluation. In fact, everything, including promotion to a higher academic title and something very practical such as monetary reward, was associated with those students' evaluations. Lily admitted that because of such pressure, she had to yield to students' preference. Her compromise was to stop or lessen speaking activities in the "Listening and Speaking" class.

## Discussion and Implications

### *Disjunction between Policymakers and Implementers*

In this study, although policymakers encouraged teachers to follow the intended curriculum with, for example, a working knowledge of syllabus, learner-centred approach, and entire English instruction, the classroom reality often left teachers unable to do so. From classroom observations and interviews, I identified a discrepancy between policymakers' intentions and teachers' actual execution of the policies. The gap existed primarily in the following aspects: understanding of the syllabus, the learner-centred approach promoted by the syllabus and textbooks, the use of English in teaching, and the impact of tests on implementation.

### *Understanding of the syllabus*

Policymakers emphasized the importance of teachers' working knowledge and thorough understanding of the syllabus in the expected implementation. However, the two interviewed teachers argued that the syllabus should have been more explicit, concrete, and specific. The results revealed that when policymakers fail to make their intentions clear, some dangers exist: either that teachers may have no clear idea of what was intended and then could ignore some aspects of the innovation or that teachers misunderstand the intentions and react with disfavour.

For example, the college English teaching syllabus failed to prescribe what teaching methods that language teachers should use. As was pointed out, the teaching methodology

14

suggested in the syllabus document is flexible (Wang & Han, 2002), "eclectic" (Fan, 1999), and "composite" (Cowan et al., 1979). From the policymakers' standpoint, the intention was to offer teachers enough freedom and space to explore or create Chinese ways of language teaching in classrooms. Policymakers preferred teachers absorbing from different approaches rather than blindly following one certain foreign teaching method. They also suggested that teachers employ flexible and practical methods according to different learners at different teaching stages.

However, from the teachers' point of view, such eclecticism implied that they could adopt whatever methods they preferred. This absence of guidance, in fact, may have encouraged teachers to stick to the teaching method they felt most comfortable with, albeit not necessarily effective or appropriate. When asked about the teaching methodology employed, the two teachers interviewed responded that they did not use any fixed teaching methods. My observations revealed that both teachers used mainly a grammar-translation method. For instance, one teacher in her "Listening and Speaking" class even asked her students to translate sentences from Chinese to English to ensure that students fully mastered the structure and its meaning. To a certain degree, the use of the grammar-translation method was counterproductive—not promoting students' communicative skills, especially speaking skill, as prescribed in the syllabus.

### *The learner-centred approach and the use of English*

Policymakers emphasized that to implement the textbook designers' teaching principles, teachers must adopt a learner-centred approach and ought to use English entirely in instruction. In contrast, my classroom observations of the two teachers confirmed that college English teaching remained teacher-centred with more teacher talk and less student talk. These findings are consistent with researchers who also elaborated that tertiary teaching and learning in China was teacher-centred, grammar-focused, and test-oriented (Wang & Han, 2002), emphasizing structure, grammar and translation, text analysis, vocabulary, rote memorization, and reproduction of knowledge (Anderson, 1993; Campbell & Zhao, 1993; Cowan et al., 1979; Lee, 2000; Li, 1984; Penner, 1995; Wang, 1999; Zou, 1998).

Previous empirical studies in the EFL context also found that a mismatch existed between policymakers and implementers regarding learner-centred approach. O'Sullivan's (2002) longitudinal case study revealed that teachers in Namibian primary schools did not implement

the learner-centred English teaching reform imported from Western communicative approaches and advocated by Namibian policymakers. Teachers' non-implementation of the reform was caused by policymakers' failure to take into consideration the classroom realities, where teachers' professional and linguistic capacity, learner capacity, support services, and poor communication functioned as hindering elements. Similarly, Karavas-Doukas (1996) examined teachers' attitudes towards the communicative learner-centred approach in Greek public secondary schools. She found that although most teachers held favourable attitudes towards such innovation, these teachers still conducted teaching in the teacher-fronted and grammar-oriented manner, an apparent discrepancy between policymakers' intentions and teachers' executions.

To further explain why such a discrepancy existed and how to bridge the gap, O'Sullivan (2002) and Karavas-Doukas (1996) said that policymakers need to bear in mind that teachers' role in curriculum implementation cannot be undermined and their classroom realities need to be adequately considered. From my interviews with the two teachers, the implementation reality was found to be the reason for this discrepancy as well. They said that because Xi'an was situated in a less developed region and student origin was mainly from rural areas, English education was not as good as coastal cities such as Shanghai or Guangdong. For this reason, teachers felt that 100% use of English and no use of the mother tongue were neither feasible nor effective for English language teaching and learning in a context such as Xi'an.

Although policymakers contended that teachers' change of perceptions was more important than their language proficiency, both teachers whom I interviewed and observed expressed their disagreement. They said that their classroom reality, especially their students' language proficiency, was the biggest concern for them to decide how much English and Chinese could be used in teaching. One hundred percent use of English in instruction would probably result in students' frustration based on their language ability. They argued that proper use of their first language was beneficial for their students: saving time, clearly conveying the course content, and more importantly, checking if students understand their instruction or not. Excluding Chinese in classrooms seemed neither realistic nor possible for them. In addition, these two teachers commented that teachers at top universities in Beijing or Shanghai undoubtedly had more of an advantage over average universities such as theirs, with top students who are more proficient on the national university entrance examinations in English. These students would have less difficulty in following teachers' English instruction in multi-media classrooms and computer

labs. Their students in Xi'an, in comparison, mainly from rural areas with lower language proficiency, struggled with instruction in English.

### *The impact of tests*

The topic of greatest difference was the use of the standardized high-stakes college English test results in evaluating teachers' performance. Policymakers reiterated that the CET was voluntary, that it should bring impact positively on classroom teaching, and that its sole purpose was to assess if the syllabus was properly implemented. The teachers revealed not only that these tests were mandatory for their universities and students, but also that tests like CET impacted negatively on them, a phenomenon discussed by Alderson and Wall (1993) and Smith (1991b).

When I observed the classroom teaching, I found the CET washback effects obvious. Although my observations deliberately focused only on the two teachers teaching freshmen, I still saw the impact of the test. Instead of following the teaching syllabus, teachers instructed on what was tested in the CET. For example, one teacher continuously reminded her students to pay special attention to words or verbal phrases such as "access," "classify," "deprive of" that were frequently tested in the CET. The other teacher spent two teaching periods explaining how to write an application letter, because the CET had such a test format. Such "teaching to the test" behaviour, with an "undesirable narrowing of the curriculum" (Alderson & Wall, 1993, p.118) has been shown as having an impact on classroom teaching and learning (Alderson & Wall, 1993; Madaus & Kellaghan, 1992; Prodromou, 1995; Smith, 1991a).

Indeed, teachers in my study conducted test-related practices that included telling students test-taking tips, emphasizing exercises potentially to be tested, demonstrating marking procedures, and using various means to promote student motivation for the test, all types of practices that Madaus and Kellaghan (1992) described. Teachers also aligned instruction with test formats and adapted their teaching to remain consistent with test contents. Specifically, the classroom observations convincingly revealed the negative washback both overtly and covertly as Prodomou (1995) delineated. The two teachers were found using examples from textbooks that emphasized the skills used in the CET. As a result, reading was given much more emphasis in the classroom than listening, writing, and speaking. The same teachers were also seen "teaching a textbook as if it were a *testbook*" (Prodomou, 1995, p. 15, original italics). Because of this, they focused on those points which would be tested, and helped students in completing

multiple-choice exercises in students' practice books. Smith (1991b) summarized that testing considerably reduces learning time, narrows the curriculum, and discourages teachers from attempting to meet goals or use materials that are not compatible with formats used by test makers. As she put it, "multiple choice testing leads to multiple choice teaching" (p. 10).

When asked in the interviews why they conducted teaching that way, both teachers expressed their quandary. On the one hand, they described such teaching as generating problematic consequences, in the fact that "even when students can get high scores and good grades in the examinations, seldom do they write fluently or speak fluently." On the other hand, they were obliged to proceed with such practices, because students evaluated and complained about them if they failed to attach importance to the CET in their teaching. Their concerns resonated with what Madaus and Kellaghan (1992) asserted about the impact of high-stakes tests on teachers. These two authors stated that the influence of testing on teachers' classroom practices is intensified when students, teachers, administrators, or the general public perceive the testing results to be linked to important decisions. These decisions include "graduation, promotion or placement of students, evaluation or rewarding of teachers or administrators, allocation of resources to schools or school districts, and school or school-system certification" (p. 139). As well, Alderson and Wall (1993) argued, "for teachers, the fear of poor results, and the associated guilt, shame, or embarrassment, might lead to the desire for their pupil to achieve high scores in whatever way seems possible" (p.118). Therefore, given the importance of this high-stakes test, it can be assumed that teachers will be highly likely to continue with the test related instruction as long as the status quo exists.

### *Implications for Implementers*

The findings of my study revealed a mismatch of intentions and executions between policymakers and implementers. Teachers as implementers did not carry out the intended curriculum. On the contrary, they conducted classroom instruction based on the context and reality where they were teaching. In Chinese college English teaching, teachers, being at the lower level of the hierarchical structure, seem to have little input into curriculum development. However, just as Connelly and Clandinin (1988) claimed, curriculum development is ultimately about teacher development, and it is teachers who decide whether implementation can be executed as it is intended by policymakers and further sustained. Therefore, Chinese EFL

teachers need to take initiative to have their voices heard about issues of language teaching and learning through various channels: conferences, workshops, and staff meetings. They should not consider themselves as passive implementers of curriculum, whose responsibility it is to simply follow the rules. Rather, they should see themselves as the major players in putting the proposed curriculum into classroom practice by actively participating in the curriculum development. They should acknowledge that their perspectives are valuable, and should have more weight in determining what and how curriculum can be implemented. Without their sincere participation, the continuity and sustainability of any innovative curriculum will be out of the question.

In addition, teachers need to be actively involved in teacher training and professional development programs such as expert seminars, academic conferences, and research undertakings. Teaching is an ongoing and life-long learning process. Teachers, therefore, should see curriculum innovation as an opportunity which facilitates them in upgrading their professional capacity. Only after they vigorously engage themselves in such ventures can they be able to actually perform ownership of the innovation. Teachers will see themselves as part of the innovation and the innovation as part of themselves as well. This ownership in turn creates more autonomous teaching instead of merely keeping to the designated textbooks. Moreover, teachers should also start seeking collaboration and collegial support from their peers. Such cooperation will eventually benefit teachers in information sharing, in conducting research, and in problem solving.

**Conclusions**

The findings revealed that between policymakers and implementers a gap seems to exist. National language policies exerted a certain impact on teachers' classroom behaviour, but teachers followed more concrete regulations from their own universities. These rules had more impact on their teaching in that disobeying them might rule out teachers' practical benefits such as monetary reward or academic promotions. In addition, the findings indicated that classroom teachers carried out instruction based on their classroom reality by primarily following first the testing syllabus and then, the textbooks. They could be said to have only partially implemented the intended curriculum.

This study reinforces the point that implementation is a very complex matter (Fullan & Stiegelbauer, 1991; Snyder, Bolin, & Zumwalt, 1992). Any coherent language curriculum will

have to reconcile a tension between desirable policy and acceptable and possible practice in context (Johnson, 1989). During the process, teachers as implementers determine whether or not curriculum innovation is executed as it is intended by policymakers in the classroom (Carless, 1999; Clandinin & Connelly, 1992; Connelly & Clandinin, 1988). Marsh (1986) said, "It cannot be expected that teachers will automatically implement a curriculum according to the intentions of the developer" (p. 19). Carless (1999) also asserted that "teachers are the individuals who implement, adapt, reject, or ignore curriculum innovation. It is thus something of a truism that they are the core of the innovation process" (p. 374). Therefore, teachers' decisive role in the implementation of the curriculum cannot be underestimated. Without their support and sincere involvement in the innovation, any curriculum implementation will stay at a superficial level, with either semi-implementation or even non-implementation.

**Acknowledgements**

**Notes**

1. To investigate such a complicated issue as the implementation of the English as a foreign language curriculum policies, the researcher explored the issue from three groups of stakeholders. Interviews with the national policymakers from the syllabus, textbook, and testing teams (referred to as policymakers in this paper) were conducted to examine the intended curriculum, interviews with the heads of the college English department were conducted to look at the administrators' perceptions, and questionnaire surveys on EFL teachers were administered to scrutinize teachers' perceptions. All the data collected from these sources were triangulated and analyzed to gain a more complete picture of the issue and findings were reported in different scholarly papers. The current study only focused on the classroom observations and interviews with the two teachers to see how they interpreted the national and institutional policies and how they implemented these policies in their classrooms.

2. This study examined the implementation of the 1999 version of the College English Teaching Syllabus and the data was collected in the summer of 2004. During the data collection, the latest version of the teaching syllabus called "College English Curriculum Requirements" (for trial implementation) was published officially by the Higher Education Department of the Ministry of Education. Since the implementation phase of this curricular document was too short, the researcher focused on the 1999 Syllabus for the exploration of the issue.

3. This abbreviation and others in this section means: TE for teacher; IN for interview; 01SU for first small university; 04 for page number in interview transcripts; 02LU for second large university.

### References

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129.

Anderson, J. (1993). Is a communicative approach practical for teaching English in China? Pros and cons. *System, 21*(4), 471-480.

Campbell, K. P., & Zhao, Y. (1993). The dilemma of English language instruction in the People's Republic of China. *TESOL Journal, 2*(4), 4-6.

Carless, D. (1999). Factors affecting classroom implementation: Task-based curriculum renewal in Hong Kong. *International Journal of Educational Reform, 8*(4), 374-382.

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 363-401). New York: Macmillan.

Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis, 12*(3), 233-239.

College English Syllabus Revision Team (1986). *College English teaching syllabus* (For students of arts and sciences). Shanghai: Shanghai Foreign Language Education Press.

College English Syllabus Revision Team (1999). *College English teaching syllabus* (For regular college students). Shanghai: Shanghai Foreign Language Education Press & Higher Education Press.

Connelly, F. M., & Clandinin, D. J. (1988). Recovery of curriculum meaning. In F. M. Connelly & D. J. Clandinin, *Teachers as curriculum planners* (pp. 81-112). Toronto: OISE Press.

Cowan, J. R., Light, R. L., Mathews, B. E., & Tucker, G. R. (1979). English teaching in China: A recent survey. *TESOL Quarterly, 13*(4), 465-482.

Denzin, N. K. (1994). The art and politics of interpretation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 500-515). Thousand Oaks, CA: Sage Publications.

Fan, C. R. (1999). Waiyu iiaoxue zhongde zhezhong zhuyi [Eclecticism in foreign languages teaching]. *Waiyu jiaoxue yu yanjiu Foreign Language Teaching and Research, 2,* 29-34.

Fröhlich, M., Spada, N., & Allen, P. (1985). Differences in the communicative orientation of L2 classrooms. *TESOL Quarterly, 19,* 27-56.

Fullan, M. G., & Stiegelbauer, S. (1991). *The new meaning of educational change* (2nd ed.). New York: Teachers College Press.

Gross, N. C., Giacquinta, J. B., & Bernstein, M. (1971). *Implementing organizational innovations: A sociological analysis of planned educational change*. New York: Basic Books.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed.). Boston: Pearson/Allyn and Bacon.

Johnson, R. K. (1989). A decision-making framework for the coherent language curriculum. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 1-23). Cambridge: Cambridge University Press.

Karavas-Doukas, E. (1996). Using attitude scales to investigate teachers' attitudes to the communicative approach. *English Language Teaching Journal, 50*(3), 187-198.

Lee, T. H. C. (2000). *Education in traditional China: A history*. Leiden, the Netherlands: Koninklijke Brill NV.

Li, X. J. (1984). In defence of the communicative approach. *English Language Teaching Journal*, *38*(1), 2-13.

Madaus, G. F., & Kellaghan, T. (1992). Curriculum evaluation and assessment. In P. W. Jackson (Ed.), *Handbooks of research on curriculum* (pp. 119-154). New York: Macmillan.

Marsh, C. J. (1986). Curriculum implementation: An analysis of Australian research studies – 1973-83. *Curriculum Perspectives, 6*(1), 11-21.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171-178.

O'Sullivan, M. C. (2002). Reform implementation and the realities within which teachers work: A Namibian case study. *Compare, 32*(2), 219-237.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.

Penner, J. (1995). Change and conflict: Introduction of the communicative approach in China. *TESL Canada Journal, 12*(2), 1-17.

Prodromou, L. (1995). The backwash effect: From testing to teaching. *English Language Teaching Journal, 49*(1), 13-25.

Silver, R. E., & Skuja-Steele, R. (2005). Priorities in English language education policy and classroom implementation. *Language Policy, 4*(1), 107-128.

Smit, B. (2005). Teachers, local knowledge, and policy implementation: A qualitative policy-practice inquiry. *Education and Urban Society, 37*(3), 292-306.

Smith, M. L. (1991a). Meanings of test preparation. *American Educational Research Journal, 28*(3), 521-542.

Smith, M. L. (1991b). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402-435). New York: Macmillan.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387-431.

Wang, H., & Cheng, L. (2005). The impact of curriculum innovation on the cultures of teaching. *Asian EFL Journal, 7*(4), 7-32.

Wang, H., & Han, H. (2002). The development of college English curriculum in China. *Carleton Papers in Applied Language Studies, 16/17,* 75-97.

Wang, H. (2001). *Integrating grammar instruction with communicative approaches for EFL students in China*. Unpublished master's research essay, Carleton University, Ottawa, Ontario, Canada.

Wang, Y. Q. (1999). "College English" in China. *English Today, 15*(1), 45-51.

Zou, Y. Y. (1998). English training for professionals in China: Introducing a successful EFL training program. *System, 26*(2), 235-248.

**Appendix A: Classroom Observation Guide**

General information:

Instructor: _____

Observer: _____

Date: _____

Time: _____

Classroom: _____

Focus of the observation:

1. Who is conducting most of the talking? Is it the teacher or the student?
2. How does the teacher teach generally? Using what methods?
3. How does the teacher interact with students? Is there any interaction between the teacher and students?
4. How does the teacher ask and respond to the students' questions in class?
5. What language activities does the teacher organize in the classroom teaching?
6. Is there any group work or pair work in the classroom teaching and learning?
7. Is there any translation exercise in the classroom teaching?
8. Is there any oral English practice in the classroom teaching?
9. How is English used in Reading/Writing and Listening/Speaking classes?
10. How much is English used in Reading/Writing and Listening/Speaking classes?
11. How is Chinese used in Reading/Writing and Listening/Speaking classes?
12. How much is Chinese used in Reading/Writing and Listening/Speaking classes?


**Appendix B: Interview Protocols for the Teachers**

Before and during the classroom observation:

1. Can you tell me something about your language learning experience?
2. Can you say something about yourself, such as your educational background and teaching experience?

After the classroom observation:

3. How would you describe your experience following the College English Teaching Syllabus?

4. What do you think of the clarity of the national College English Teaching Syllabus? Is it clear and easy to follow? If not, why not?

5. What do you think of the practicality of the Syllabus in terms of your language teaching? Is it practical for you to use? If not, why not?

6. How appropriate is the College English Teaching Syllabus for your teaching context?

7. Why do you think that teachers should have a national College English Teaching Syllabus?

8. To what extent do you think the textbooks accurately represent the College English Teaching Syllabus? If not, where are they lacking?

9. What is the benefit of the College English Test from your perspective as a teacher?

10. What impact, if any, does the College English Test have on your classroom teaching?

11. What impact, if any, does the College English Test have on your students' learning?

12. How do you actually conduct your classroom teaching to achieve the objectives stated in the official syllabus?

13. How much support do you receive from your department head regarding how you teach College English?

14. What kinds of support would you like to receive from your department head in your teaching of College English?

15. What are the most challenging aspects that you face regarding implementing the College English Curriculum?

16. What is the rationale for you to conduct your class in English / in Chinese?

17. Why do you organize your classroom activities that allow your students to work with their peers / to work alone?

# Testing Oral Language Proficiency of University EFL Students

**A. Majid Hayati**
*Shahid Chamran University of Ahvaz, Iran*

**Ehsan Askari**
*Azad University of Masjed Soleiman, Iran*

**A. Majid Hayati**, an Associate Professor of Linguistics, holds a doctorate degree in Linguistics from the University of Newcastle, Australia. He teaches TEFL, Language Testing, Linguistics and Contrastive Analysis at Shahid Chamran University of Ahvaz. Hayati has published a number of articles in Roshd Magazine (Iran), Reading Matrix (USA), PSiCL (Poland), Asian EFL Journal (Korea), and Arts and Humanities in Higher Education (England).

**Ehsan Askari** received his MA in TEFL from Islamic Azad University of Ahvaz in 2006. He has been teaching English at Azad University of Masjed Soleiman, Iran since 2003. His special fields of interest are Testing English as a Second/Foreign Language, Linguistics and Translation.

**Abstract**
The present study aimed at developing a series of objective criteria for measuring and scoring the oral proficiency of EFL students in moving toward a more objective mode for scoring the oral language proficiency. To achieve this purpose, eighty students from the University of Masjed Soleyman in Iran were selected based on their availability and their successful passing of conversations one, two, and three. Then, their oral proficiencies were rated against a validated and newly-developed checklist. The obtained scores were compared with the group's performance in their previous conversation courses. Result indicated a low correlation between the two groups of scores. It was also proved that the subjective measures were not reliable enough to indicate the students' abilities in terms of oral language proficiency**.**

**Key words:** Oral Language Proficiency, Objective Scores, Subjective Scores, Scoring Criteria.

**Introduction**

Many language tests follow a psychological rather than linguistic theoretical framework, evidenced by the use of a single modality (such as a paper-and-pencil test that ignores spoken and oral comprehension) (Pray, 2005). Most current tests of oral proficiency have the same

deficiencies, and many of the measures used by the teachers share the problem of subjectivity. This status is sustained by factors such as large classes, teachers' inadequate command of English, and the lack of easy access to support materials and facilities (Ramanathan, 2008, Sook, 2003). Therefore, due to the complicated nature of this skill, testers and language teachers should make use of reliable analyses for the purpose of objectivity.

The focus of the present study is on the fact that in university conversation classes there exists no clear-cut checklist or a hard and fast set of criteria for measuring the oral proficiency of students majoring in English. Various types of tests designed and administered- mostly paper and pencil listening tests, student-student, and student-teacher interviews rated without using established criteria- are not suitable to the mode. Therefore, an objective and integrated checklist is needed to measure the students' competence on the basis of their performance. To do so, the researchers appropriately modified the existing checklists to include an important factor, "communication" which is essential for the purpose of assessing levels of oral ability, to help the test designers move from subjective teacher-made tests towards a more standardized testing of oral/aural skills. This checklist was developed as comprehensively as possible so that the researchers were able to take into account most of the required criteria in the tests for measuring oral proficiency. Sample models for developing this checklist were extracted from Farhady, Jafarpur, and Birjandi (2001), Heaton (1990), Hughes (2003), IELTS Testing Center (2000), and Underhill (1987). The most significant criteria considered in the checklist included accent, speed of response, diction, listening comprehension, communication, and fluency to name but a few.

Until now, several studies have been conducted in developing measures for evaluating language learners' oral proficiency. Harris (1968) suggests a list of criteria for measuring oral skills, which is technically known as "Sample Oral English Rating Sheet". Harris's sample comprises five criteria to be rated: pronunciation, grammar, vocabulary, fluency, and comprehension, each of which includes 5 levels. The proficiency guidelines for speaking were developed in 1982 by the American Council on the Teaching of Foreign Languages (ACTFL) with the purpose of creating a criterion that could be used to identify the foreign language proficiency of speakers ranging from "no knowledge" of EFL to "total mastery" gained through widespread application. The ACTFL guidelines include: superior, advanced (high, mid, low), intermediate (high, mid, low), and novice (high, mid, low) levels.

Also, Underhill (1987) has offered a rating scale for measuring speaking skills. A rating scale,

as defined by Underhill includes 1) very limited personal conversation, 2) personal and limited social conversation, 3) basic competence for social and travel use, 4) elementary professional competence, and 5) general proficiency of all familiar and common topics.

One area of decision-making in rating scales is scoring. Farhady, Jafarpur, and Birjandi (2001) state that depending on the objective of the a test, scoring may be done holistically or discretely; the former refers to an overall impression according to which the interviewee either receives excellent, good, fair, or pass/fail scores. The latter, on the other hand, rates the interviewee's performance separately on scales that relate to accent, structure, vocabulary, fluency, and comprehension. Another crucial work in this realm is a checklist developed by Hughes (2003). The checklist assigns the candidates (interviewees) to a level holistically and rates them on the six-point scale of each of: accent, grammar, vocabulary, fluency, and comprehension. The test is both given and rated by the teacher with no student self-evaluation and self-judgment about their progress.

However, more recent studies, emphasizing the interactional aspect of language, have focused on learners' awareness of the test procedures. For example, a different view of language assessment, inspired by the idea of Task-Based Instruction (TBI) is casting light on the field of foreign language testing. In task-based language assessment (TBLA) language use is observed in settings that are more realistic and complex than in discrete skills assessments, and which typically require the integration of topical, social and/or pragmatic knowledge along with knowledge of the formal elements of language (Mislevy, Steinberg & Almond, 2002). In another case, Lambert (2003), giving the tests at the end of term to nine classes of between 26-31 first year Japanese university students majoring in electrical and mechanical engineering, predominantly male, upper elementary to pre-intermediate level, concludes that recordings of the student-student interviews would provide a clear justification for the marks awarded and it is also a good idea to give the students a chance to think about what they would say by putting the actual test roles on the Intranet. In the light of the above studies, it should be clearly noted that the current scoring methods applied in Iranian universities are mostly impressionistic, based on experience and lack validity and reliability; the checklist proposed can be utilized as an alternative method in order to obtain objective scores which are true representative of the students' oral communicative ability. So, the current study could function as a prerequisite to interactional approaches to language testing since its main goal is to suggest a rather valid and

reliable checklist as a measurement device for assessing oral proficiency. In other words, the same checklist could be used by both teachers and students in methods such as TBLA, student-student interviews, etc.

**Questions of the Study**

For the sake of arriving at an objective decision, this study pursued to provide answers to the following questions.

1. Which measure, subjective or objective, provides a more valid and reliable estimate of the oral proficiency of the EFL learners?

2. Is there a meaningful relationship between the subjective and objective sets of scores?

**Methodology**

*Participants*

Subjects in the present study were 80 students selected from the students of English Language Teaching at Islamic Azad University of Masjed Soleyman. The rationale for their selection was their availability and the fact that the participants had already passed three conversation courses successfully and they were also taking conversation four at the time of the study. Twenty five percent of the participants were male (n=20) and the rest were female (n=60), ranging from 20 to 27 years old.

*Instrument*

One instrument utilized in the process of the present study was the proposed checklist including a series of standards and criteria for measuring oral communicative abilities of EFL students on an academic level. Another instrument was the IELTS format of interview (a speaking test) in which the interviewees were asked to answer general and personal questions about their homes and families, jobs, studies, interests, and a range of similar topic areas in about five minutes**.** The other instrument utilized in the present study was a tape recorder for recording the interviews.

**Procedure**

In order to validate the newly designed checklist, that is, to determine the extent which the checklist measures what it is supposed to measure, a pilot study was conducted. Ten students

were randomly selected and rated using both the new checklist and the one designed by Hughes (2003) to determine the criterion-related validity of the new checklist. The correlation coefficient obtained between the two series was 0.968 indicating that the new checklist was valid.

By the end of the semester, the subjects were asked to speak for about two minutes on a particular topic for which they were given almost two minutes to think about. All of the selected subjects were interviewed and rated against the new checklist by two raters, first by one of the researchers and then by a bilingual (a native speaker of English who speaks Persian). Each interview session was held in the presence of one of the researchers, the classroom teacher (both as the interviewers) and one of the subjects (as the interviewee). Each interview commenced with a set of simple questions and then proceeded to more challenging ones, and before each session, the subjects were asked to explain and write down brief notes on the sources and textbooks which they had practiced in conversation courses and also the methods applied during the courses and the final examinations. This was done as a warm up activity to decrease the psychological stress and to ensure that the same mode and channel had been used to score the oral proficiency of the subjects in the previous courses. In order to enhance the reliability of the scores, rating activities were carried out first by one of the researchers and then by an inter-rater, and agreement was reached on each student's score.

All the subjects' scores in conversation courses were collected from the Educational Affairs Department of Masjed Soleyman University, and their average scores were calculated. After gathering the required data, the next step was to rate and score each interview based upon the developed checklist with the aim of attaining more reliable and objective scores. The correlation coefficient determined whether or not there was a possible relationship amongst these series of scores.

**Data Analysis**

The interviews obtained during the study were assessed through listening to the recordings, and the performance of each interviewee was rated on the basis of the criteria indicated in the developed checklist first by the researchers and next by an inter-rater. After calculating the average of the scores given by two raters, two series of scores were attained-the average scores in interview and the average scores in conversation courses.

By using the Microsoft Excel software (2003 version) and calculating the variables, the

correlation obtained was 0.0045 which indicated that the correlation between the two series of scores was substantially low. This proved the hypothesis of the study that the previous ratings were wholly implemented in a subjective manner compared to the ratings made against the newly developed checklist including the objective criteria.

Minimum and maximum values were higher for the subjective rating than for the objective ones which might indicate that instructors in conversation courses were more generous, and so these scores do not represent the true oral abilities of the subjects. Also the mean of the students' scores in conversation courses was 15.87 while the corresponding mean score in interview was 11.77. However, the difference between the standard deviation of both groups was not meaningful and indicated that the use of standard criteria for scoring oral proficiency caused the scores of the students to fall off in a similar manner, i.e. the subjects who received higher scores among others by subjective scoring measures also received the higher range of scores by the objective measures although their range of scores lowered meaningfully in objective scoring. The median of the scores in conversation courses was 16 which showed that half of the scores were higher and half of them were lower than 16. The median of the objective scores was 12.

Table 1 illustrates the average scores of conversation courses assigned to students by their instructors through traditional subjective means of testing and scoring oral communicative abilities. 70 percent of the scores were in the range of 15 to 20, and the rest of the scores fluctuated between 12 and 15.

Table 1. Descriptive statistics on objective   and subjective measures

| *Statistical Evaluations* | *Subjective Scores* | *Objective Scores* |
|---|---|---|
| Population | 80 | 80 |
| Min. Value | 12.33 | 8.17 |
| Max. Value | 19.33 | 15.67 |
| Range | 7 | 7.5 |
| Mean | 15.87 | 11.77 |
| Standard Deviation | 1.70 | 1.70 |
| Standard Error | 0.19 | 0.19 |
| Median | 16 | 12 |
| Sum | 1,270.1 | 942.38 |
| Sum of Squares | 20,395.27 | 11,331.72 |
| Variance | 2.92 | 2.92 |

Figure 1 represents the average scores given by the two raters to the same groups of students

based on the standard criteria listed in the designed checklist (See Appendix for a sample checklist). The distribution of these scores was lower than the course scores with 50 percent of the scores between 12 and 16 and the rest between 8 and 12.

Figure 1 *The average scores of subjects based on the standard checklist*
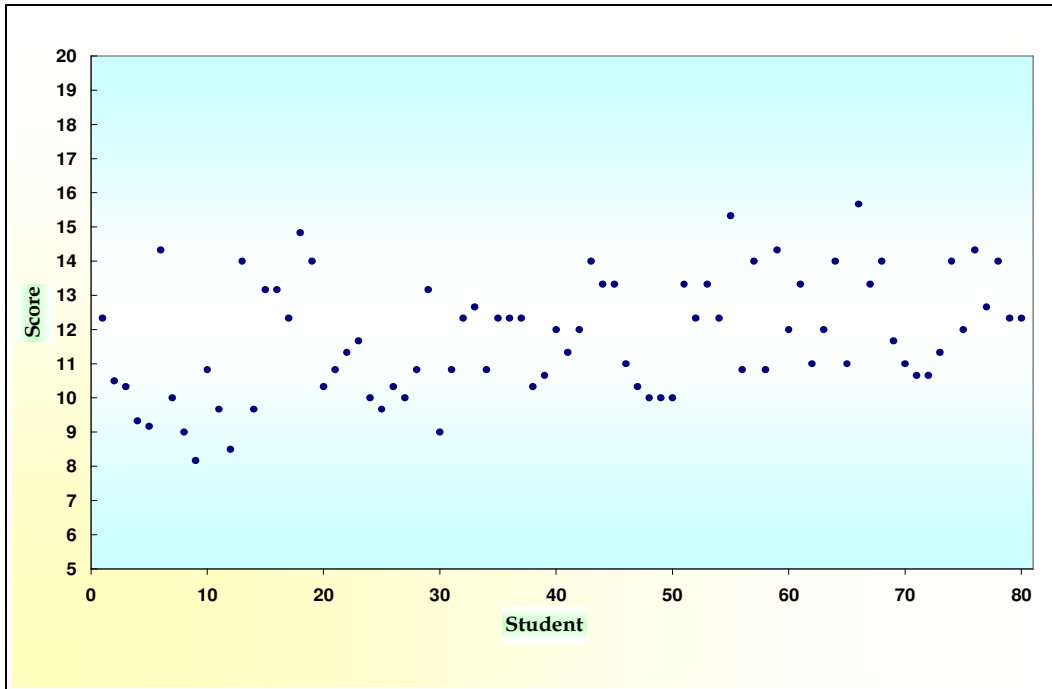


This distribution of scores indicating the variation in the students' oral abilities shows that the actual abilities of the students are far below that obtained by their EFL teachers.

In order to determine the contribution of each scale on the objective scores and the performance of the subjects, the scores in various scales were specified in terms of six scales.. Although the general performance of the subjects was weak, the figure shows the strength and weaknesses of the subjects in different sub skills of the speaking skill.

The checklist contains 6 scales namely, fluency, comprehension, communication, vocabulary, structure and accent, each of which includes 5 levels of proficiency. The performance of the subjects on each scale was then independently calculated.

Figure 2 *The percentages of scales of the checklist contributed to the total score*
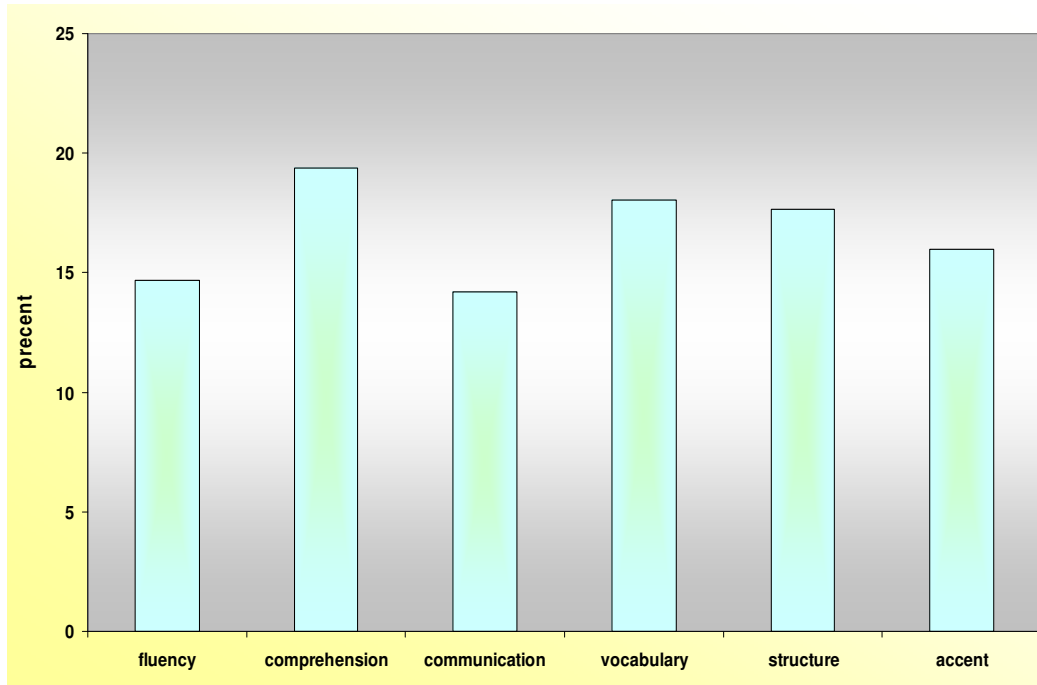


The performance of the subjects in the areas of comprehension, vocabulary, and structure was fairly better compared to that of fluency, communication, and accent (Figure 2).

**Discussion**

One point to discuss here is that teachers' scoring the students' oral proficiency subjectively is neither reliable nor valid, and so the given scores cannot present the true ability of the subjects in oral language proficiency. By analyzing oral language proficiency in terms of a number of scales and calculating the learners' ability in terms of their performance on the scales, the researchers could now validly judge the learners' oral language proficiency. The general performance of the subjects, however, was weak in the ratings carried out, but their performance in the individual scales of the checklist was varied. That is, in certain scales they performed successfully but in others they did not.

Results showed that the performance of the subjects in linguistic components was better than their performance in communicative aspects. Fluency is one of the key factors in assessing the oral language proficiency. Most of the subjects in the present study were hesitant and their oral

performances were discontinuous. Another scale on the checklist was comprehension in which the subjects showed a better performance than in the other scales. In most cases they understood the question or/and the gist but were not able enough to manage the discussion. We suggest that the comprehension skill of the subjects should be assigned a higher priority in the development of the English teaching curricula.

On the scale of communication, the subjects had the weakest performance indicating the greater attention they need to pay to this aspect of their communicative competence. Although the performance of the subjects in vocabulary and grammar scales was better, there were other problems such as lack of complete accuracy that should be considered by the EFL teachers. As for acceptable and intelligible accent, interviewees showed a weak performance in this scale which may be indicative of the EFL learners' ignorance of this part of language.

Analyses showed that the mean score of the subjects in objective scoring was approximately four points lower than the mean score in subjective scoring. It might be that subjective scoring was implemented based on personal judgments and also the scores were allotted to the overall speaking skill of the subjects, and therefore, the range of scores was high. On the other hand, in the objective scoring, in light of the standards and criteria, the communication skill as a whole was broken into six distinct sub-skills. The scores obtained for each sub-skill were summed up in order to represent the total score given to each subject in terms of their comprehension ability, and so, the range of scores was meaningfully lower. Through objective scoring, weak and strong sub-skills of the subjects' speaking skill can be assessed enabling TEFL teachers to remove the deficiencies and reinforce the stronger points.

**Conclusions**

In sum, the point to be taken into account is the lack of attention and application of specific standards to score learners' oral productive skill. In the EFL setting, there are many teachers who score the learners' speaking ability subjectively without applying any criteria and they often show generosity in scoring; consequently, the obtained results will be a series of unreliable and invalid scores which are not truly representative of the learners' actual ability. (However, there may be few language teachers who, after a long time of experience, use their intuition to score the learners' performances subjectively. They are an exception though.) Therefore, in order to obtain better results including more reliable and objective scores in testing speaking, it is

essential to utilize a series of criteria to score oral language proficiency. As Pray (2005) mentions, "Oral-language assessments must measure the essential elements of knowing a language, not just lexical knowledge. This includes the ability to produce new utterances and recombine forms to represent ideas, events, and objects on an abstract level, to produce forms of the language they have never heard before, and to demonstrate mastery over the general functions of language such as syntax, morphology, semantics, and pragmatics" (p.405).

One concerns of teachers is how to prepare reliable tests for measuring oral proficiency of the students and score their performance. To have a more reliable estimate of the students' oral language ability, using a checklist will be very helpful. It will eliminate all those sources that threaten the stability of the test scores. The checklist can act as a blueprint to teachers who wish to assess their students' oral proficiency. It reminds them of the macro-skills as well as the specifications or micro-skills that should be included in testing oral proficiency.

**Delimitations**

Despite the promising results, this study suffered from a few problems. One shortcoming was related to our population which was predominantly female; the results of the present study, therefore, might not be generalizable to the male population. Moreover, speaking skills, though emphasized, are overshadowed by other skills due to lack of environment to adequately practice or apply oral/aural skills. This results in a series of problems especially while conducting the interviews during the research.

**References**

American Council on the Teaching of Foreign Languages. (1982). *ACTFL provisional proficiency guidelines*. Yonkers: ACTFL.

Farhady, H., Jafarpur, A., & Birjandi, P. (2001). *Testing language skill: From theory to practice* (9th ed.). Tehran: SAMT Publication.

Harris, D. P. (1986). *Testing English as a second language*. New York: McGraw Hill.

Heaton, J. B. (1990). *Writing English language tests* (3rd ed.). New York: Longman.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Lambert, I. (2003). Recording speaking tests for oral assessment. *The Internet TESL Journal, IX*(4). Retrieved 23 March 2008 from http://iteslj.org/Articles/Lambert-SpeakingTests.html.

Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing, 24*(4), 489–515.

Mislevy, R.J., Steinberg , L.S., Almond, R.G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*(4), 477-496.

Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal*, *29*(2), 387-409.

Ramanathan, H. (2008). Testing of English in India: A developing concept. *Language Testing, 25*(1), 111-126.

Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal,* Retrieved 23 March 2008 from file:///C|/asian-efl-journal/dec_03_sub.gl.htm.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques.* Cambridge: Cambridge University Press.

**Appendix**

**The Sample Checklist for Measuring Communicative Abilities:**

Extracted from Farhady et al. (2001), Harris (1968), and Hughes (2003)

**<u>Scale I- Fluency:</u>**

☐ 5- Speaks fluently.

☐ 4- Speaks with near-native like fluency, pauses and hesitations do not
    interfere with comprehension

☐ 3- Speaks with occasional hesitations.

☐ 2- Speaks hesitantly and slowly because of rephrasing and searching for words.

☐ 1- Speaks in single word and short patterns, unable to make connected sentences.


**<u>Scale II- Comprehension:</u>**

☐ 5- Understands academic discourse without difficulty.

☐ 4- Understands most spoken language except for very colloquial speech.

☐ 3- Understands academic discourse with repetitions, rephrasing, and clarification.

☐ 2- Understands simple sentences, words; requires repetitions, slower than normal speech.

☐ 1- Understands very little or no English.


**<u>Scale III- Communication:</u>**

☐ 5- Communicates competently in social academic settings.

☐ 4- Speaks fluently in a social academic setting, errors do not interfere with meaning.

☐ 3- Initiates and sustains conversation, exhibits self- confidence in social situations.

☐ 2- Begins to communicate for personal and survival needs.

☐ 1- Almost unable to communicate.

### Scale IV- Vocabulary:

☐ 5- Uses extensive vocabulary in any domain appropriately.

☐ 4- Uses varied vocabulary to discuss general topics and in special interests.

☐ 3- Uses academic vocabulary, some word usage inappropriate, slightly damages the

message.

☐ 2- Uses limited vocabulary, constant use of one word.

☐ 1- Inadequate basic vocabulary.

### Scale V- Structure:

☐ 5- Masters a variety of grammatical structures, almost no error.

☐ 4- Occasional grammatical errors but no problem with understanding.

☐ 3- Uses some complex sentences but lacks control over irregular forms.

☐ 2- Uses predominantly present tense verbs, constant errors interfere with understanding.

☐ 1- Severe errors make understanding completely impossible.

### Scale VI- Accent:

☐ 5- Acceptable pronunciation, with few traces of foreign accent.

☐ 4- Speaks with few phonemic errors, but almost intelligible pronunciation.

☐ 3- Occasional errors necessitate attentive listening.

☐ 2- Constant phonemic errors make understanding extremely hard.

☐ 1- Severe problems make understanding almost impossible.