Theoretical Basis of Communicative Language Assessment of Speaking Skills

Dr. Shamim Ali,
Assistant Professor,
Department of English & Applied Linguistics,
Allama Iqbal Open University,
Islamabad,
Pakistan

Abstract

Teaching and Testing are compatible partners; both contribute to the success of any language program. This study aims to assist teachers to understand the philosophy of oral exams, and this, consecutively will help them to identify effective strategies for the development of competency and the performance of students. It would have a valuable effect on the learning & teaching process. From the students point of view it will motivate them to learn and to develop more ability in speaking skills. The focus of this research is twofold, firstly to discuss the theoretical basis of communicative language learning & testing, secondly to assess the speaking skills of learners at the end of the session, Bachman's assessment tools (1990) Reliability and inter-rater reliability provided me an instrument to measure my students' Speaking skills, therefore the focus of my study is the theory of communicative competence. The idea of this study is to analyze the technique, process, and outcomes of the oral assessment rubric in NUML and to check reliability and inter-rater reliability of the assessment tools. Ten learners from Diploma and Foundation classes were tested and evaluated on the basis of their communicative performance. Four competent Raters evaluated them in accordance with the judgment tools designed by the language teachers. Outcome showed strong correlations for Diploma levels and poor to strong correlation for Foundation levels and strong correlations for Diploma levels. The implication of this research confirmed the need to revise the appraisal process and at the same time to provide proper training to language teachers towards a proper assessment task.

Key Words/Terms: Communicative Language, assessment, oral skills.

I. Introduction

Assessment of speaking skills is quite different from other assessments as it requires a very systematic study. Teaching is mostly followed by testing and hence we come to know not only about the progress that students have made but also about the aims and objectives we set for the students, the efficacy of teaching material that we used while teaching, the effectiveness of teaching methodology that we used and evaluation of the assessment itself. To some extent it can be considered very challenging as compared to other skills, because all language teachers have laid down different criteria for performance assessment of their students. Trained teachers usually start from the beginning of the session, which is called formative assessment, whereas the teachers who do not have any prior training just apply summative assessment at the end of the session. Teaching Assessment and evaluation are common concerns of all language teachers. There is difference between Assessment and Evaluation. Assessment gathers students performance over a period of time whereas Evaluation takes place when the marks are awarded after the completion of any activity, task, test or classroom quiz. Now a days very few teachers are fully aware of the utility of both assessment & Evaluation, but good institutions still facilitate their teachers to employ it on regular basis. As this study is based on the theoretical basis of communicative language testing, but this study will only highlight the procedure of assessment. With the introduction of CLT in language teaching the speaking skills has become very important component, so assessment has also become one of the most significant element in language teaching. Assessment can be considered as a scientific or a philosophical study. It is based on our daily routine, but it has not been given much attention even the concept of assessment is not clear to our
teachers and policy makers. Therefore our learners are assessed on an annual basis, and learners can predict beforehand about the exam. However it cannot be considered as an achievement of the students and the self-evaluation of teaching methodologies. Teachers do not have proper training for these kinds of assessment and in the long run it has a great impact on assessment; and reliability of the assessment can be answerable. As I have mentioned earlier that assessment procedure is very systematic and consistent however the results can be inconsistent because of time, setting and style of Rater. The evaluation of all raters can cause a substantial amount of disparity in measurement which cannot be ignored, because the assessment style of each rater varies. Rating of speaking skills is at a level of great intensity, and the first encounter with the rater is very impressionistic, therefore it is important to train language learners in accordance with communicative approach. Assessment of Speaking skills is considered as a power of orderly thought, because examiner/rater logically assess distinct from any kind of predisposition. It is necessary to define all aspects of assessment techniques to avoid or to manage discrepancy. This is natural when evaluating aspects communicative effectiveness, fluency and language ability hard to define, can be troublesome. In the result Subjective assessments need to be checked for reliability. The high reliability elements that contribute are (1) assessment tools that have clear and complete judgment criteria (2) explicit definition for each level (3) clear testing for each level of teaching, for monitoring learners’ progress (4) teacher training programs should be catered. (5) Teachers should join on mutual consensus. National University of Modern Languages is considered an ideal university for data collection. Students come from varied cultural and linguistic backgrounds, and this is the only university that teaches all four skills along with the component of Error Analysis in language classes to second language learners. Last year in 2010 I decided to investigate the practices of speaking classes, teaching methodologies, teachers and their beliefs. For the sake of this study the teachers were asked about their experience as language teachers, the handouts or reading materials for developing speaking skills, components of spoken languages, frequency check and reasons for evaluation/assessment. The outcome validated that most of the speaking skills teachers don’t employ formative assessment in the beginning of the course and they have designed rubrics for summative assessment only. These teachers do not have proper training in the field of testing & assessment of speaking skills, secondly these teachers are not graduated from NUML. Whereas the teachers of NUML are trained in that manner because a very important subject TEFL is being taught in their final semester, and they have done several sessions of micro teaching. The unawareness, in terms of language teaching, of teachers from the other universities is due to the ignorance of TEFL subject. As speaking skills is the most important component of language learning programme so it should be instructed in proper manner. The purpose of speaking skills is to enhance the confidence level of students and to make them aware of their thinking skills. Therefore the teachers should be trained in developing purposeful methods for instruction, assessment and evaluation and they should become experts in the field of teaching and testing of speaking skills.

Our language planners, syllabus designers, and curriculum designers can be blamed for not developing a proper rationale for our language teachers. These people still follow obsolete teaching and assessment methodologies that are rooted in grammar translation methodology and teacher’s centre class room. The criteria to assess speaking skills are unclear and biased, and it lacks consistency. Teachers do not have proper collaboration and it leads them to inconsistent assessments. The implication of this study was the dire need for tutorials in the field of assessment and evaluation for speaking skills teachers the overall improvement of the system of evaluation and assessment, so the system should have consistency in assessment practices (Munoz, et al., 2003). Considering all these repercussions, I took initiative discuss it with my colleagues and designed rubrics for assessing our students. These rubrics were based on the communicative competency and the theoretical model by (Canale and Swain, 1980; Savignon, 1983; Bachman, 1990 was chosen. Swain himself was in the favor that all kinds of tests should be built on a theoretical framework. For that purpose I selected the outline of Canale and Swain (1980) model. In order to have a comprehensive model of communicative competence, Bachman’s (1990) and Shohamy’s (1988) models were also not neglected by me. The academic environment at NUML is deemed optimal for language learning as well as for testing as it is the only university where more than 30 languages are taught using communicative
methodology. As the requirement of communicative methodology is to check the competence and to impart knowledge in all four strands therefore all learning and teaching is based on four skills along with Error Analysis, which serves as grammar component of language learning & teaching. Therefore the purpose of oral test formats is to gauge the competence and performance of the learners through interaction, negotiation and meaningful communication. Below are the descriptions of oral test formats including direct tests and semi-direct tests. Direct tests consist of simulation techniques, group discussions, picture-stimuli procedures, and oral interviews. Taped exams and semi-direct test are viable alternative for those teachers who do not get an opportunity to administer direct tests. These tape exams are also helpful to assess the oral exam later (Larson, 1984). Certainly, these test formats are not comprehensive. It is hoped that more test formats will be put forth by teachers. Therefore, the rubric was planned on linguistic, strategic, discourse, sociolinguistic abilities, simulation techniques, group discussions, picture-stimuli procedures, and oral interviews, and to measure the effectiveness of this rubric an investigation was done.

2.Literature Review
2.1 System Of Assessment
(Rating) The rating system can be considered as a vital part in assessment and it should be integrated into the process of test design. There are two approaches to the assess the performance of Speaking Skills. a) Holistic Assessment b) Analytic Assessment. Both of these have pros and cons, holistic scoring is done on the basis of an overall performance, it can be the first impression or the overall impression as well. Raters or Assessor who are very well experienced, and they gauge the students performance immediately. The benefit of using the holistic approach is that it is very less time consuming. Raters or Assessors score immediately, but this type of assessment is only beneficial when the Assessor or Rater is experienced, secondly if we have large class for assessment. Analytic scoring, on the other hand, focuses on specific aspects such as Grammar(Tenses-Structure), Fluency, Strategies (Cognitive Metacognitive), Sociolinguistic competence, Vocabulary items and pronunciation (Omagi, 1987). There are a number of advantages to analytic rating. First it gives teachers a breakdown of each student's language proficiency, which can be used for diagnostic purposes (Jons, 1986). Secondly, it compels teachers to consider aspects of performance which they might otherwise ignore (Hughes, 1989). However, rating by the analytic approach will take more time than rating by the holistic approach. Another question regarding the analytic approach is whether the sum of the part equals the whole? According to Cliford (1989), the answer is no. The main reason is that "not all of the individual factors contributing to oral language proficiency may have been identified". To put it differently, the listed aspects are not comprehensive enough to account for oral language proficiency. Former criteria for assessment was based on the production of language. Recognizable mechanism of language proficiency commonly used were: Tenses, Structure grammar, Vocabulary Items, Pronunciation, and Subject Matter. Whereas, the recent criteria of assessment is based on the Functional aspects of language. It identifies a number of elements: Extensiveness (how long is the speech produced?) Complexity (how far does the speaker knit intricate web of words?) Pace (what is the rate or apace of the speech fast does he or she speak?) Suppleness (can the speaker adapt quickly or is he flexible?) Precision (Is the use of Language accurate?) Appropriacy (The choice of vocabulary is suitable) Organizatation (does the speaker bring about his thought independently or is he self governed?) Reiteration (Does he/she want probes or stimulus?) Reluctance (Does the speaker hesitate before and while speaking?) Some of these elements, such as extensiveness, complexity, and pace, can be assessed objectively, whereas others like appropriacy precision, suppleness and Origination, are not very easy to judge. These elements of assessment can be judged by trained and experienced raters. Reliability and validity are two significant qualities to gauge the test scores, the purpose is to establish the effectiveness of a test. The validity checks the errors which cause variation in language test scores, whereas the Reliability is characterized as consistency of measurement. In other words, it is the degree to which test scores for a group of test takers are reliable over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker" (Berkowits, Fitch, and Kopriva, 2007). There are three ways to calculate and...
check reliability, each of which check different kinds of inaccuracies and errors. No. 1 is Internal consistency (split-half reliability) it measures within the test and scoring procedures. No. 2 is stability (test-retest reliability) it calculates consistent test scores are over time. No. 3 is equivalence (alternate form) it indicates the extent to which points on alternate forms of a test are the same. The reliability of language tests may be negotiated if they these tests have been given qualitative assessment instead of quantitative assessment (Weir 1990). Prior to proceeding to the marking stage raters coordination is important for awarding of marks: there should be coherence and consistency in the distribution of marking. To get sufficiently high rater reliability for test results the proper training of the raters is essential. A small tutorial before the summative assessment should be held for promoting proper learning and agreement among raters. These kinds of tutorials would help the examiner to comprehend the philosophy of rating scales to toil with, and it would also make them able to interpret their descriptors consistently (Alderson & Wall, 2001). Green (2002) also emphasized on the need of shared understanding between the raters to avoid the variance in the manner of interpretation. How ever the assessment should be unbiased and free of distortion Brown (2001) considered Validity & reliability two important tools for measuring the consistency of assessment, these tools also analyze predisposition, clarity and alteration of assessment. Reliability in the first place also explains the consistency of assessment therefore the rater should be well trained in measuring the internal consistency of the items. In other words the rater should be completely qualified of checking reliability, for example, if they are gauging or measuring a quiz and test takers obtains an item correct, at the same time he or she will also get other, similar items correct. Rater can assess twice free from inaccurate stereotypes. At this stage if the raters are experienced and they possess complete knowledge of assessment then they would check the reliability which is correlation between scores and it refers to the consistency in measurement. If they find the reliability of a test above 80, they will consider it very good reliability; if it is below 50, it would not be considered a very reliable test at this stage raters can have alternate ways. Inter-rater reliability was the most important aspect, and all the raters were very much thoughtful about it. Increasing this type of reliability calls for some practice on the teachers’ part and an appraisal tool where lucid performance criterion is specified. At the same time Validity refers to the precision of an assessment for that purpose validity of the test can be measured in three ways, and these three ways we decide on the test scores all three kinds, content, criterion and construct of validity evidence should be considered.

3. Research Questions
Q.1. what is the best method to assess the performance of Second Language learners’ English speaking ability until that learner reaches a level at which s/he can take a proficiency test in English?
Q.2. What kinds of test accommodations does NUML provide for its Diploma and Certificate learners?
Q.3. What language tests Rater at NUML likely to involve for assessment?

4. Method Participants/Students
Ten students chosen at random from the department of English (Functional Courses) Five learners from Foundation Level levels, and Five learners from Diploma Level.

4.1. Raters
Four subject specialists of English were chosen in assessing English using the rubric designed by me. Designated Rater one was an experienced language teacher with an experience of more than 10 years.
years, she did her TESOL certification from abroad. The second Rater had a language teaching experience of fourteen years; the third rater had an experience of ten years; this rater had done her masters in linguistics from the same University and her expertise was in Phonetics & Phonology. The last rater was also a Language teacher with an experience of six years.

4.2 Instruments

A rubric and five degree rating scale was designed to assess the learners speaking skills performance. Five major components were chosen for assessing the student’s. To measure the verbal expression the Communicative Effectiveness was chosen, for clarity grammatical aspects were chosen, for checking their level of comprehension, task completion and vocabulary were chosen, and finally for checking their phonology the element of pronunciation was also added in the rubric. The rating scale was selected between 1 to 5. Five was to be considered the highest mark, and three marks were regarded as the minimum passing grade. The details of major components are mentioned below and assessment sheet is appended in the Appendix section. Communicative Effectiveness: Providing them a relaxed vibe with deliver a message to check the progress of speech. It also gauges students' capability to make use of different strategies, initiation and maintaining speech process. To compose changes for communication failure and to begin and keep hold of speech construction. Features to keep in mind: Pausing/ Hesitation, in direct way of saying something, Imitating, coining of words, explanation, describing and telling summary. Simulation Techniques Jones (1985) defines simulation as "the process of abstracting some aspect of reality and concretely representing it in the form of a specific task that examinees are expected to perform" In using imitations, a teacher sets up tasks with the combination real contextual variables, for that purpose and the test taker is expected to perform a given task according to the specifications set by the teachers. Perhaps the best known language testing simulation technique is role-playing. The tester describes a situation and asks the test taker to play a part in it. The role plays used for the test may vary from short simple role plays involving only two or three students to far longer role plays involving ten to twenty students. Jones (1985) suggests that the tester must be skilled in controlling the situation while at the same time making it appear as real as possible. Heaton (1988) suggests that, in addition to problem-solving activities and puzzles, tasks involving consensus-seeking are particularly suitable for group discussion. The members of the group are given a particular situation and instructed to make various decisions. It is necessary for them to use the target language to justify their decisions and seek agreement from the other members of the group. Group discussion offers pedagogical as well as scheduling advantage (Mueller, 1987; Reves, 1982). From the pedagogic standpoint, it is interactive testing at its best. Interaction among students increases rapidly as students respond to the ideas and comments offered by other students in discussions. Furthermore, group oral testing appears to be an alternative when individual oral testing presents scheduling problems. However, oral group testing might be inappropriate for students who are shy and introspective, therefore the best methods for their assessment can be the communicative Effectiveness, grammar structures syntax, recognition of consonants and production of sounds. At this stage the accent should not be given priority, but simple Vocabulary: in terms of functions of language. Pictures, maps, and diagrams can be used in oral language testing. There are several techniques to elicit the oral language by using the picture-stimuli procedure. The most common way is for learner to be given a picture to study for a few minutes; they are then required to describe the picture in a given time. Weir (1988) states that it can be an efficient procedures and one of the few available to get the test taker to provide an extended sample of connected speech. It is, he claims, also useful for gauging his competence to use particular tenses such as the past tense for reporting. Although it does not tell much about the competence of test taker, but an experienced rater can assess competence through his performance. However there are several methods to assess test takers ability to interact orally. Another way of using the picture-stimuli procedure is to ask the test taker questions based on the pictures given. This method is more highly recommended due to the greater involvement in mutual communication. The questions asked need to have possibilities for elicitation of meaningful speech in a connected context of discourse where a chronological and cause-effect type of relationship is obtained between a series of events. Green (2002) stressed upon the need of thorough assessment to overcome the problems of inaccurate judgments. He further states that in order to overcome the
problems of inaccurate judgment, it is necessary to develop a communal understanding. among raters, because it forms positive influence and leads to accurate assessment for that reason communal understanding and comprehensive meeting are needed before the conduction of speaking skills test.

4.3 Procedure
As I have mentioned earlier that a shared understanding was needed, therefore the four raters held several standardization meetings before test. After discussions, some adjustments to the rubric with proper and mutual understanding the raters proceeded to assess learners individually. All learners were called for interview on two different days: Foundation Level, on the First day Diploma level on the Second day. Foundation level interviews lasted from 15 to 30 minutes, based on different kinds of activities. Exactly in the same manner diploma class interviewees were interviewed. The evaluators encouraged the test taker to respond each questions and speak about their personal life, choices, ambitions and dislikes. Relatively the questions for both levels were same.

5. Result & Data Analysis
Inter-rater reliability was worked out through mean score comparison and correlations among the four assessments of Raters. Mean score comparisons were performed through analysis of variance (One Way ANOVA). Pearson Correlation was established at 0.05 to check significant differences statistically between the means of the 4 raters, because the P-values were higher than 0.05.

<table>
<thead>
<tr>
<th>One Way Anova p-Values</th>
<th>Communicative Effectiveness</th>
<th>Pronunciation</th>
<th>Grammar</th>
<th>Vocabulary</th>
<th>Task Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of Four Raters. Assessment per Aspect (Levels 5, 1)</td>
<td>0.4213</td>
<td>0.7319</td>
<td>0.7783</td>
<td>0.7031</td>
<td>0.8661</td>
</tr>
<tr>
<td></td>
<td>0.5921</td>
<td>0.6286</td>
<td>0.8114</td>
<td>0.6861</td>
<td>0.7815</td>
</tr>
</tbody>
</table>

The table shows a summary of Pearson product moment correlations range among the 4 Raters for the five aspects on which rubric was designed. It measured the strength of the linear relationship between the variables. These correlation coefficients range between (−1) and (+1) summarizes the range of correlations for the different aspects for levels 5-10. The result shows high correlation among all the evaluators, but there was not a significant difference statistically found among the raters' mean scores, the correlations for the different five aspects range from poor to moderate being the lowest, Grammar and Communicative Effectiveness and the highest Task Completion, Vocabulary, and Pronunciation, very weak in grammar and vocabulary. The study of one Way Anova for levels 5 – 10, together with all aspects and all raters become evidence for that there are no statistically significant differences among the raters' mean scores.
Table Two: Range of Correlations

<table>
<thead>
<tr>
<th>Assessment of Four Raters.</th>
<th>Correlations range (levels 1-4)</th>
<th>Assessment per Aspect</th>
<th>Range of correlations ( levels 5 -10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td>Correlation range</td>
<td>Aspect</td>
<td>Correlation range</td>
</tr>
<tr>
<td>Communicative Effectiveness</td>
<td>(-0.2361 – 0.5718)</td>
<td>Communicative Effectiveness</td>
<td>(0.4900 -0.8093)</td>
</tr>
<tr>
<td>Grammar</td>
<td>-0.2361 – 0.5718</td>
<td>Grammar</td>
<td>(0.3325 -0.2066)</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>(0.1042 – 0.4554) low</td>
<td>Pronunciation</td>
<td>0.4773 -0.6310</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>(0.2337–0.6745) low</td>
<td>Vocabulary</td>
<td>(0.3327 -0.6422)</td>
</tr>
<tr>
<td>Task Completion</td>
<td>(0.3344 – 0.5337) low</td>
<td>Task Completion</td>
<td>(0.4699 -0.7664)</td>
</tr>
</tbody>
</table>

Anova test indicated that there was not a significant difference statistically found among the raters’ mean scores, the correlations for the different five aspects range from poor to moderate being the lowest, Grammar and Communicative Effectiveness and the highest Task Completion, Vocabulary, and Pronunciation, very weak in grammar and vocabulary. One Way Anova analyses for levels 5 – 10, including all aspects and all evaluators shows that there are no statistically significant differences among the evaluators’ mean scores.

6. Discussion

Correlation was found at fundamental levels (1-4) ranging from poor to moderate. Inter rater reliabilities for Grammar, Communicative Effectiveness and vocabulary were checked on different basis. First, grammar revealed the lowest correlations of all features. There was a great variation found in the assessment of raters. It indicated that all raters perceived this skill differently because they did not develop a shared understanding before summative assessment. Secondly it was found that a criterion of assessment set for Foundation level was very high. It was a need for coordination between raters for good assessment. Data revealed that all raters did not follow any discussion before the assessment took place and all of them did not set any criteria to judge mutually. It was also revealed that the (achievement standards) for Foundation levels were not devised according to their level. The exercises in the rubric placed great emphasis on the accuracy of structures. There was a need to place emphasis on students’ ability to comprehend the message and to convey the message properly. Performance standards were set very high for Foundation students in Communicative Effectiveness, whereas for Diploma learners it was accurate; that is why diploma level learners performed well in it. Communicative Effectiveness requires students to be able to negotiate meaning and employ strategies to compensate for disruption in utterances. These kinds of compensation were quite obvious in Diploma level students. Strong correlations were given for Task Completion. To a great extent, Task Completion is influenced by the successful realization of the other features included in the rubric. If learners are able to complete the task using wrong English then there is clearly something wrong with the task or its administration. Pronunciation and vocabulary reliabilities were moderate and showed tendency to be higher as the rubric became more demanding for higher levels. On the other hand, the correlations obtained from levels 5-10 provided evidence of inter- rater reliability. The strength of the relationship showed that the different language features could be judged consistently by all the evaluators who interpreted the rubric descriptors reliably, thus maximizing objectivity during assessment. Employing different category of raters brought dynamism to the process of assessment. Furthermore, the precision of the rubric for Diploma levels allowed...
raters to judge properly among different abilities and award suitable marks. The highest award was indicated for Communicative Effectiveness of Diploma level learners. A strong consensus on Communicative Effectiveness was indicated. NUML places great emphasis on the functional nature of the syllabus which relies on communication activities, such as the book for Error –Analysis in their grammar class and BBC Course Books. It helps learners to speak English in real life contexts.

7. Conclusion

Following the theoretical model by Swain (1985) and Bachman (1990), a frame work was designed to assess students and to provide wash back for teachers. The principle of wash back would help teachers to assess themselves as well. The teachers should always significantly regard this theory, that the test will have a great impact on their teaching methodology/ability as well, so the test’s influence is positive. This kind of research would help them in transforming their teaching of speaking skills, because speaking skills is an ability to convey thoughts and ideas orally in a way that others understand and to understand what others say as well so as to respond adequately. As Carroll (1986) puts it, we should “look for tests which vitalize teaching, not lay a dead hand on it” the ability both to convey thoughts and ideas orally in a way that others understand and to understand what others say is important.

The data revealed that NUML has laid down a significantly high proficiency standard for its Foundation learners. It is important to devise a new, appropriate rubric for Foundation Students. At the same time, this type of test provided an evidence to appraise Diploma Level Learners; hence it may be concluded that is necessary to establish distinctive parameters for different levels. If, for instance, we set up requirements which second language learners are unable to meet, they may lose confidence and then give up studying the language. This study can contribute much to our teachers’ understanding for designing a valid speaking test according to the student’s level. All speaking skills teachers should know the logic behind these scales. They should know how to interpret the scale properly, they should determine the criteria before the assessment task takes place, for that purpose considerable training is required for teachers for attaining adequate reliability levels. It is the need of time. There should be a proper training of the language teachers as formal raters. Language teachers will be in the better position of assessing the performance and abilities of their students. The rationale behind this kind of assessment was not categorizing students or to challenge language teachers, but to understand learners’ strengths and weaknesses so that instruction can be attuned to learning outcomes. It is sought-after that through the use of proficiency-oriented oral assessment, constructive wash back effects will appear in classroom teaching and learning. Ideally, language classrooms teachers will not only focus on grammar but rather work on second language in communicative situations. Consequently, learners will allocate themselves for developing oral language ability in the Second language in addition for the improvement of the other three skills, i.e., Listening, Reading (Receptive Skills) and Writing (Productive Skills).

References

4. Bachman, L. (1991) ‘What does language testing have to offer?’, TESOL Quarterly, 25, 671-


8. Carroll, B. J. (1985). Second language performance testing for university and professional context. In P. C. Hauptman, R. Leblanc, & M. B. Wesche (Eds.), Second language performance testing (pp. 73-88). Ottawa:


