ASSESSING SPEAKING SKILL
Dera Estuarso

A. SPEAKING SKILL

1. Definition of Speaking

Speaking is the real-time, productive, aural/oral skill (Bailey, 2003:48). It is real time because the other interlocutor is waiting for the speaker to speak right then and the speaker cannot revise his response as he might do in writing. It is productive because the language is directed outward. It is aural because the response is interrelated with the input often received aurally and it is oral because the speech is produced orally.

2. Levels of Speaking

From the highest to its lowest level, speaking can be dissected into text, utterance, clause, phrase, word, morpheme and phoneme (van Lier, 1996). Success in speaking means being able to communicate message using accurate and acceptable use of language throughout these levels. Knowing these levels shall help test maker understand what to expect from test taker's performance.

3. Types of spoken Language

Spoken Language can be in the form of monologue or dialogue. A monologue can be planned or impromptu while dialogue is almost always unplanned while dialogue can be interpersonal or transactional; each can be either familiar or unfamiliar.

4. Micro- and Macroskills of Speaking

Brown (2004:142-143) suggests a list of micro- and macroskills of speaking to help determine test maker as what to assess (whether to assess on smaller chunks of language or speaking's larger elements) as follows:

Microskills
1. Produce differences among English phonemes and allophonic variants.
2. Produce chunks of language of different lengths.
3. Produce English stress patterns, words in stressed and unstressed positions, rhythmic structure, and intonational contours,
4. Produce reduced Forms of words and phrases_
5. Use an adequate number of lexical units (words) in order to accomplish pragmatic purposes.
6. Produce fluent speech different rates of delivery.
7. Monitor one's own oral production and use various strategic devices—pauses, fillers, self-corrections, backtracking—to enhance the clarity of the message.
8. Use grammatical word classes (nouns, verbs, etc.), systems (e.g., tense, agreement, pluralization), word order, patterns, rules, forms.

9. Produce speech in natural constituents— in appropriate phrases, pause groups, breath groups, and sentences
10. Express a particular meaning in different grammatical forms.
11. Use cohesive devices in spoken discourse.

Macroskills

12. Appropriately accomplish communicative functions according to situations, participants and goals.
13. Use appropriate styles, registers, implicature, redundancies, pragmatic conventions, conversation rules, floor –keeping and –yielding, interrupting, and other sociolinguistic features in face to face conversations.
14. Convey links and connections between events and communicate such relations as focal and peripheral ideas, events and feelings, new information and given information, generalization and exemplification.
15. Convey facial features, kinesics, body language, and other nonverbal cues along with verbal language
16. Develop and use a battery of speaking strategies, such as emphasizing key words, rephrasing, providing a context for interpreting the meaning of words, appealing for help, and accurately assessing how well your interlocutor is understanding you.

B. ASSESSING SPEAKING

1. Challenges in Assessing Speaking

Hughes (1984:101) believes that that successful interaction involves both comprehension and production. For that reason, he believes it is essential that a task elicit behavior (or performance) which actually represent the test taker's speaking competence. In addition to selecting the appropriate assessment, O'Malley (1996:58) also mention determining evaluation criteria as another major challenge. Much in the same tone, Brown (2004:140) describes two major challenges in assessing speaking: (1) the interaction of listening and speaking (e.g. the use of much clarification) can make it difficult to treat speaking apart, (2) the speaker's strategy to dodge certain form to convey meaning may make it difficult for test makers to design a solid elicitation technique (one that can result in the expected target form).

2. Basic Types of Speaking Assessment Tasks

Brown (2004:141) provides 5 types of Assessment Tasks. The headings below are Brown's proposed categories but the tasks in each category come also from the descriptions by Heaton (1988), Hughes (1989) and O'Malley (1996). In the past it is agreed that speaking leaves no tangible product to be assessed (unlike writing), yet today technology has make it possible to record the speech in every type of the task. A

challenge of this sort has little relevance to today's practice. Therefore, albeit unmentioned, the following types of Task may involve recording the test taker's speech.

a. **Imitative**: repeating a small stretch of language and focused on pronunciation.

Test maker considers using this type of assessment if he is not interested in test taker's competence in understanding and conveying meaning or in getting involved in interactive conversation. The competence assessed is that of purely phonetic, prosodic, lexical and grammatical (pronunciation).

b. **Intensive**

1) Reading Aloud

Heaton (1988:89) and Hughes (1989:110) maintains that the use of reading aloud may not be appropriate because of the difference in processing written input from that of spoken one. However, a check on stress-pattern, rhythm and pronunciation alone may be conducted using reading aloud. Brown (2004:149) suggests that we use reading aloud as a companion for other more communicative tasks.

2) Directed Response Task (e.g response to a recorded speech)

One of the most popular Task of speaking for its practicality and mass lab-use, despite its mechanical and non-communicative nature, DRT is beneficial to elicit a specific grammatical form or a transformation of a sentence which requires minimal processing (microskills 1-5, 8 & 10) (Brown, 2004:147).

3) Sentence/Dialogue Completion

Heaton (1988:92) warns us the fact that this type may provide illogical flow of conversation given that the sentence or dialogue completion is done in lab (which is what normally administered). Therefore, this type will probably be beneficial only for assessing test taker's microskill of providing the right chunks of language and other pronunciation features.

However, as Brown (2004:151) exemplifies, a more responsive-type of sentence/dialogue completion may actually be free of said caveat and keep us from the risk of judging a test taker's competence as insufficient caused by aural misunderstanding in processing the input. SDC helps measure speaking competence apart from its interrelatedness to listening.

4) **Translation** up to simple sentence level (interpreting-game)

Interpreting, as Hughes (1989:108) describes, may involve the test-proctor acting as native speaker of test taker's first language and the test taker interpreting the utterance into English. It is believed that because speaking is negotiation of intended meaning (O'Malley, 1996:59), interpreting-game can be used to measure test-taker competence in conveying his message into the target language (Brown, 2004:159).

5) **Limited picture-cued Task** (including simple sequence)

Pictures are mostly convenient to elicit description (Hughes, 1989:107). In addition to describing comparison, order of events, positions and location, a more detailed picture may be used to elicit test taker's competence in telling a plan, directions and even opinions (Brown, 2004:151-158).

c. **Responsive**:

Small dialogue, response to spoken prompt (simple greeting, request & comments)

1) **Question and Answer**

Questions at responsive level tend to be referential (as opposed to intensive, display question) (Brown, 2004:159). Referential question requires test takers to produce meaningful language in response. Such questions may require an open-ended response or a counter-question directed to the interviewer (Brown, 2004:160).

2) **Giving Instruction and Direction**

In this type of task, test takers are elicited their performance in describing a how-to description. A five- to six-sentence response may be sufficient to be required either from an impromptu question or a-minute planning prior to the instruction (Brown, 2004: 161).

3) **Paraphrasing**

Oral Paraphrasing can have written or aural input with the latter being more preferable. A paraphrase as a speaking assessment should be conducted with caution because test taker's competence may be mistakenly judged by their short-term memory and listening comprehension instead of their speaking production.

d. Interactive (larger dialogue on Transactional and Interactional Conversation)

1) Interview

Interview can be face-to-face, one-on-one or two-on-one each with its advantage and disadvantage. A two-on-one interview may save time and scheduling and provide authentic interaction between two test takers, although it pose a risk of one test taker domination the other.

Hughes (1989:105) proposes 11 rules to conduct an interview:

1) Make the oral test as long as feasible
2) Include as wide a sample of specified content as is possible in the time available
3) Plan the test carefully
4) Give the candidate as many 'fresh start' as possible
5) Select interviewers carefully and train them
6) Use a second tester
7) Set only a tasks and topics that would be expected to cause candidates no difficulty in their own language
8) Carry out the interview in a quiet room with good acoustics
9) Put candidates at their ease
10) Collect enough relevant information
11) Do not talk to much (the interviewer)

In addition to Hughes' proposal, Canale (1984) proposes four main steps to follow to conduct, in this case, an oral proficiency test.

1) Warm Up   : small talk about identity, origin and the like
2) Level-Check   :*wh*-questions, narrative without interruption, read a passage aloud, tells how to make or do something, a brief guided role-play
3) Probe   :field-related questions
4) Wind-down   : easier questions pertaining test taker's feeling about the interview

The challenge with an interview is how the open-ended response is scored. Creating a consistent, workable scoring system to ensure reliability has been one of the major challenge in designing an interview as means to assess speaking (Brown, 2004:171). There are at least two solution to this problem: one is using an analytical scoring rubric and the other is a holistic one. Rescoring the performance later from the tape can be an alternative, too (O'Malley, 1996:79).

2) Drama-like Task

O' Malley (1996:85) divides drama-like task into three sub-types: improvisations, role play and simulation. The difference of each is respectively the preparation and scripting. Improvisation give very little opportunity for test taker to prepare the situation and may incite creativity in using the language. Role play provides slightly longer time to and test taker can prepare what to say although scripting is highly unlikely. Meanwhile, simulation (including debate) requires planning and decision making. Simulation may involve real-world sociodrama which is the pinnacle of speaking competence.

Like interview, drama-like task may evoke unpredictable response. Similar care used to tackle interview may be useful for this type of task as well.

3) Discussions and Conversations

Discussions and Conversations (Brown, 2004: 175) provide somewhat similar difficulties in terms of predictability of the response hence consistency of the scoring to that of interview and drama-like tasks. Test makers seem to choose this type of task as *informal assessment* to elicit and observe test taker's performance in:

1) starting, maintaining and ending a topic
2) getting attention, interrupting and controlling
3) clarifying, questioning and paraphrasing
4) signaling for comprehension (e.g nodding)
5) using appropriate intonation patterns
6) using kinesics, eye contact and body language
7) being polite, being formal and other sociolinguistic situation

4) Games

It is nearly impossible to list all games, but virtually all games that can elicit spoken language objectively can be used as informal assessment for speaking. Brown (2004:176) warns us that using games may go beyond assessment and adds that a certain perspective need to be maintained in order to keep it in line with assessment principles.

Some examples of games which Brown (2004:175-176) mentions (tinkertoy, crossword puzzle, information gap, predetermined direction map) can all fall in the umbrella of information-gap activities by O'Malley (1996:81)'s

standpoint as he explains that an information gap is an activity where one student is provided information that another (e.g his pair) does not know but need to. An information gap activity involves collecting complete information to restructure a building, sequence a picture into order or simply find the differences between two pictures. To score an information gap activity, O'Malley (1996:83) suggest test maker to consider the speaker's "accuracy and clarity of the description as well as on the reconstruction."

e. Extensive (monologue)

The following are monologues which take longer stretch of the language and requires extensive (multi-skills) preparations. The terms are self-explanatory and some may actually possess some characteristics with some types previously explained only with longer and broader scope of language use.

1) Speech (Oral Presentation or oral report)

It is commonly practiced to present a report, paper or design in school setting. An oral presentation can be used to assess a speaking skill holistically or analytically. However, it is best used for intermediate or advanced level of English focusing on content and delivery (Brown, 2004:179).

2) Picture-cued Story Telling

Similar to the limited version, at this level the main consideration of using a picture or a series of pictures is to make it into a stimulus for longer story or description; a six-picture sequence with enough details in the settings and character will be sufficient to test, among others, vocabulary, time relatives, past tense irregular verbs and even fluency in general (Brown, 2004:181)

3) Retelling a Story, News Event

Different from paraphrasing, retelling a story takes longer stretch of discourse with different, preferably narrative, genre. The focus is usually on meaningfulness of the relationship of events within the story, fluency and interaction to audience (Brown, 2004:182)

4) Translation (Extended Prose)

In this type of task, a longer text preferably in written form which is presented in he test taker's native language is to be studied prior to interpreting the text with ease in the actual testing. The text can cover a dialogue, procedure, complex directions, synopsis or a play script. Caution should be made concerning with this type of task because this particular type

requires a skill not intended for every speaker of a language. Therefore, if this type is to be used a degree of confidence should be made sure (as in the case whether the test takers are in pursuit of a bachelor degree!) (Brown, 2004:182).

3. **Scoring Rubric**

An effective assessment should follow this rule (Brown, 2004:179):

(1) Specific criteria

(2) Appropriate task

(3) Elicitation of optimal output

(4) Practical and reliable scoring procedures

Scoring remains the major challenge in assessment. There are at least two types of known scoring rubric for speaking: (1) holistic and (2) analytical. A holistic rubric range, for example, from 1 to 6 each reflecting unique capacity of the speaker with 6 being normally native-like traits and 1 a total misuse of language which incite misunderstanding. An analytical rubric, on the other hand, scores performance in different subcategories such as grammar, vocabulary, comprehension, fluency, pronunciation and task completion. There are two common practice regarding the latter: (1) the total score is summed in average to reflect an overall score or (2) each categories is given a different weight sometimes without the necessity to sum up the total score.

O'Malley (1996:65) suggests several steps in developing rubric:

(1) Set criteria of task success

(2) Set dimensions of language to be assessed (grammar, vocabulary, fluency, pronunciation .etc)

(3) Give appropriate weight to each dimension (if omission is possible, do)

(4) Focus of what test taker *can* do, instead of what they *cannot*.

Which rubric is better? Whichever is used, if high accuracy is the goal, multiple scoring is required (Hughes, 1989:97) Since test taker's speech can now be recorded for second-time scoring by different rater, a balance between holistic and analytical rubric (i.e use two types of rubric for the same task whenever possible) is recommended (O'Malley, 1996:66).

C. CONCLUSION

The key of assessing speaking skill is understanding the continuum of (1) spoken language, (2) task types and (3) scoring rubric. This non-rigid separation between one level of competence and another requires time and effort in specifying the criteria of speaking, task to elicit particular behavior and in developing practical yet representative scoring rubric. The variety of task types will help test maker to decide which one is appropriate for the wide array of the continuum of this particular skill.

REFERENCES

Bailey, K. M. (2003). Speaking. In D. Nunan, *Practical English Language Teaching* (pp. 47-66). Singapore: Singapore.

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices.* White Plains, NY: Pearson Education.

Heaton, J. B. (1988). *Writing English Language Tests (new edition).* London: Longman.

Hughes, A. (1989). *Testing for Language Teachers.* Cambridge: Cambridge University Press.

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic Assessment for English Language Learner: Practical Approaches for Teachers.* White Plains, NY: Addison Wesley.

van Lier, L. (1996). *Interaction in the Language Curriculum: Awareness, Autonomy and Authenticity.* London: Longman.