Communicative Language Testing
Sieu Phan

Abstract
Together with the communicative language teaching approach, there has been an increase in awareness and research on communicative tests used to measure language learners' ability to use the target language in authentic situations. In order to design tests to serve this purpose, test makers should bear in mind core principles and characteristics. A question about whether communicative test makers can make accurate measurement of test takers' language ability based on the scores or not is also addressed in this paper. Related to the use of communicative language testing, some literature reviews show that this kind of test may challenge test makers, which is discussed at the end of the paper.

Introduction
Traditionally, most language tests aim at testing knowledge about the language, such as testing knowledge about vocabulary and grammar. However, according to Brown (2003), "By the mid-1980s, the language-testing field had begun to focus on designing communicative language-testing tasks" (p. 10). This means that the need for communicative language test has been recognized, and much research on communicative language tests has been done since then. I hope that by doing this literature review, I can explore the principles and characteristics of communicative test design, factors that impact test scores beyond communicative language ability, and challenges in communicative testing.

Principles and Characteristics of Communicative Test Design
Principles
Communicative language tests are used with the goal of measuring language learners' ability to take part in acts of communication or to use language in real life situations. Communicative tests, which cover the four language skills of listening, speaking, reading, and writing, are designed on the basis of communicative competence. According to Canale and Swain (1980), communicative competence involves linguistic competence (knowledge of linguistic forms), sociolinguistic competence (the ability to use language appropriately in contexts), discourse competence (coherence and cohesion), and strategic competence (knowledge of verbal and non-verbal communicative strategies) (p. 4). Understanding the model of commu-

nicative competence is necessary and helpful for developing communicative language tests, which involves formulating test objectives and considering the effects of the test on teaching and learning. Following the model of communicative competence, a team at the Ontario Institute for Studies in Education (OISE) has established four principles of communicative test design (Bailey, 1998).

The first principle is to "start from somewhere." The study of Katsumasa (1997) and Wesche (1983) showed an agreement with this statement by confirming that when designing the test, test makers should state carefully what they expect test takers to perform when they use the target language in a specific context, which means that test writers must know what they want to test. After that, test makers can establish scales and criteria in assessment procedures to measure exactly the stated features of testees' performance Bailey (1998) provided an example to illustrate this principle, which is that even though conveying and capturing meaning while maintaining accuracy are two important elements in communication, if the tester focuses on checking the test takers' ability to convey meaning, then accuracy will not be put into the scoring criteria. Therefore, it is unfair for students if test writers take into account other unstated objectives when grading tests, which also negatively affects the test's validity.

"Concentration on content" is the second principle in designing a communicative language test. The content here refers to not only topics but also tasks that will be implemented. Appropriate content matches

or fits learners' or test takers' age, proficiency level, interests, and goals/needs. According to Carroll (1983), "the language tasks our learners are expected to perform in their future jobs will guide us with the tasks we will set them in our tests" (p. 37). In other words, tasks should be constructed based on the testee's relevant needs. For instance, some tasks for children at the end of primary school may include introducing themselves and others, reading traffic signs such as "Danger," "No Entry ," and so forth. The tasks should aim to be authentic and have clear reference in reality. These tasks match children's proficiency level and children's age. Children are supposed to be able to do the tasks because what they do is what society (at least based on opinions of experienced teachers and advisors) expects of them in real life.

The third principle of communicative test design is "bias for best." This means that test makers should bear in mind that they should create a test that can exploit test takers' performance at their best. The work of Brown (2003) also supports this principle in designing the test. According to Brown (2003), "biased for best" is "a term that goes little beyond how the student views the test to a degree of strategic involvement on the part of student and teacher in preparing for, setting up, and following the test itself" (p. 34). To illustrate this third principle, Bailey (1998) said that when she gave a test that requires students to do a dictocomp, she noticed that the text might be higher than students' proficiency level; therefore, she read the text three times and also encouraged them to ask for new words involved in the text. Swain (1984) declared that in order to create an assessment procedure which is biased for best, test makers and teachers should provide students or test takers appropriate review to help them to be well-prepared and ready for the test, suggest strategies that will be beneficial, and construct the test in a way that it is modestly challenging to the best students/testees and yet the weaker will not be overwhelmed (as cited in Brown, 2003, p. 34).

"Working on washback" is the fourth principle of communicative language testing.

In order to obtain positive washback, test writers should create clear scoring criteria that will be provided to *both* teachers and test takers. Course objectives and test content are also put into consideration in the hope of promoting positive washback. According to Wesche (1983), when formulating objectives, it is important to clarify (a) the purpose of the interaction concerning topics and the language functions that the learner will need, (b) the context that may impact language use, and (c) appropriate types of discourse, and the degree of skill at testees' levels. Clarification for these factors not only promotes positive washback but also makes it easier for test makers to choose good stimulus material that will provide necessary language forms such as structures and vocabulary.

The principles constructed by the team at the OISE are not unique. Katsumasa (1997), citing Morrow's (1981) study, discussed similar principles in the context of communicative language testing. These principles are (a) knowing what you are measuring, (b) measuring the ability to deal with discourse, (c) focusing on communication processes (d) setting up real situations, and (e) understanding the fact that mistakes are not always mistakes because minor grammatical mistakes do not frequently hinder communication and correct grammar is not the only requirement for successful communication . After reading and comparing the principles of communicative language test design put forth by Katsumasa and the OISE team, I prefer the framework of the OISE team because it involves the discussion of how to achieve beneficial washback. Hughes (2003) indicated that "backwash is now seen as a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large" (p. 53). Therefore, the OISE emphasis on trying to achieve positive washback is welcome.

Characteristics
In addition to the principles that the team at OISE has established, communicative language tests also have some core characteristics that test makers should follow to create

4

a successful test. According to Brown (2005), there are five requirements for setting up a communicative test. These include meaningful communication, authentic situation, unpredictable language input, creative language output, and integrated language skills (p. 21). First, the test needs to be based on communication that is meaningful to the students, that is, it meets their personal needs. It should promote and activate language which is useful for them. Making use of authentic situations can increase the likelihood that meaningful communication will be achieved. More importantly, "language can not be meaningful if it is devoid of context" (Weir, 1990, p. 11). Furthermore, communicative tests also offer students the opportunity to encounter and use the language receptively and productively in authentic situations to show how strong their language ability is. By putting emphasis on "unpredictable language input" and "creative language output," Brown referred to the fact that, in reality, it is usually impossible to predict what speakers will say, i.e., language input or to prepare for one's reply, i.e., language output. This natural way of communication should be replicated in a communicative test. The last characteristic of a communicative test is that it will elicit the students' use of combined language skills, as is the case in real life communication.

Tasks in Communicative Language Assessment

Based on features of communicative tests and components of communicative assessment, Brown (2005) concluded that performance assessment and task-based assessment are two appropriate ways of designing communicative tests. Both are important in language testing environment, however, in this paper, I would like to focus on task-based assessment as it was described in the work of Norris, Brown, Hudson, and Yoshioka (1998).

Brown, Hudson, Norris, and Bonk (2002) offered the following definition of task-based language assessment (as cited in Brown, 2005):

In task-based language assessment, then, we are interested in eliciting and evaluating students' abilities to accomplish particular tasks, or task types in which target language communication is essential. Such assessment is obviously performance assessment because a student's second language performance on the task is that which gets evaluated. (p. 24)

Below, I will cite some examples of communicative assessment tasks around the theme of environmental issues (these tasks are reproduced from Norris et al., 1998). In my view, these tasks can be suitable for intermediate/advanced ESL/EFL students who want to improve their communicative ability in English. Specifically, the students' goals are to be able to employ all language skills effectively, to use with accuracy certain vocabulary, and expressions used in certain domains and tasks, such as environment, social problems, and persuading people to participate in social activities using ads. These are language functions that they may encounter in their real life.

Task 1: Comparing environmental organizations

Prompt: Find the section in *Save Our Planet* that compares different environmental organizations and their efforts. Familiarize yourself with the charts in this section. Then answer the questions about environmental organizations as they are posed.

Realia/materials: *Save Our Planet* book (recent survey of environmental issues and efforts that are being taken on behalf of environmental preservation/protection-final section in the book is composed of a chart that compares major environmental organizations and the types of issues that they address); set of questions not to be seen by the examinee, rather first exposed through the test prompts (questions ask examinee to identify different organizations that address different issues and how much membership costs).

Task 2: Expressing views on the environment and pollution

Prompt: Read the two short views told by people from different countries about the environment and human pollution. Then compare, using your own words, the two views to the typical point of view in your country. Do people in your country think more like one of the two views, or do they have very different views from those expressed in the two readings?

Realia/materials: Two short points of view, told by the 'man on the street' (written as a person would speak, simple vocabulary and structures, common and accessible ideas- one point of view represents the individual as responsible for maintaining a clean environment, the other represents the innocence of the individual and the responsibility of the government and big industry to take care of thing).

Task 3: Product packaging ranks

Prompt: Examine the set of products from your local grocery store. Pay special attention to their packaging. Now create a list of the products, ranking them from the most environmentally safe packaging to the least. After creating your list, explain your top two and bottom two choices. Why did you place these products at the top and the bottom of the list?

Realia/materials: Set of products (photographs or actual products, if possible) numbering around ten (with obvious differences in degree of environmental consciousness, from the small and completely recyclable/organic to the large, superfluous packaging that is typical with many products); list with blanks for the ranking of products (from best packaging to worst packaging).

Task 4: Organize advertisements

Prompt: look through the stack of advertisements from magazines. Separate the advertisements into files. One file should have ads that show products which seem to be friendly to the environment. The other file should have ads that do not address the environment or seem unfriendly to the environment. For each ad, write a brief sentence explaining why you chose to put it in a particular file.

Realia/materials: Set of advertisements for different products (selected from different magazines- each ad either keys on some kind of environmental science, e.g., our product is dolphin sage, or does not, e.g., smoking is fun, with inclusion in one or the other category fairly obvious) totaling no more than 15; two file folders; page within each folder for short explanation of why the ad is included. (p. 133-135)

This series of test items consist of different tasks which are consistent with themes relevant to ESL/EFL learners at intermediate/advanced levels who want to use English effectively in familiar, real-life contexts. These tasks also show connections which make it easy for teachers to design follow-up activities if they are going to use these tasks in a language classroom. The designers of these tasks also included ample opportunities to assess students' performance. For example, students have to use four skills in order to complete the task: writing a list, reading short views, watching (listening to) the video, and taking notes. Test makers can evaluate students' outcomes based on the quality of their work, for instance, do they get accurate information from the video? Do they make any serious mistakes when they write the lists? Can you understand students when they give the presentation? However, these tasks may not apply to all situations; teachers will need to adapt them to fit their situations. For instance, although the book *Save Our Planet* may not be available to teachers in some countries, they can still apply the framework used in this sample test with different materials.

There are many tasks in this series and teachers may use them differently. I think it is possible to have students do all the tasks in one test; however, it is time consuming.

Teachers can separate the tasks and use them appropriately. I believe that either way will not affect the value of the task, and test makers can still measure students' communicative language ability. As test makers assess learners' performance, they need to establish scoring criteria. Fortunately, Norris et al., (1998) provided guidelines for developing complete scoring criteria for evaluating students' performance on a task. Based on the work of Herman, Aschbacher, and Winters (1992), they suggested these steps in assessment:

1. investigate how the assessed discipline defines quality performance

2. gather sample rubrics for assessing writing, speech, the arts, and so on as models t adapt for your purposes

3. gather samples of students' and experts' work that demonstrate the range of performance from ineffective to very effective

4. discuss with others the characteristics of these models that distinguish the effective ones from the ineffective ones

5. write descriptors for the important characteristics

6. gather anther sample of students' work

7. try out criteria to see if they help you to make accurate judgments about students

8. revise your criteria

9. try it again until the rubric score captures the 'quality' of the work (p. 67)

As with other kinds of tests, steps in developing the scoring criteria for tasks used to assess learners' ability to perform in the target language requires a lot of responsible and cooperative work, agreement on the application of rubric, and trials before it is applied.

Factors That Impact Test Scores
In addition to the effect of communicative language ability, other factors such as test method facets, personal attributes, and ran-

dom factors also affect strongly examinees' performance on tests for communicative ability. Bachman (1990) grouped test method facets into five categories:

1. testing environment, including familiarity of the place and equipment, personnel, time of testing, and physical condition;

2. test rubrics involving test organization, time allocation, and instruction;

3. the nature of the input, concerning format, and nature of language;

4. the nature of the expected response, referring to format, nature of language, and restrictions on response;

5. the relationships between the input and the response in language tests whether reciprocal, nonreciprocal or adaptive relations. (p. 119)

The last facet, reciprocal language use, refers to "the use of language by one individual to produce an effect in another individual through the reduction of uncertainty with knowledge of results" (Bachman, 1990, p. 149). In other words, in spoken or written interaction between people, what one person says affects what the others can or do say. Unlike reciprocal language use, nonreciprocal, which includes reading, listening to lectures, and so on, has "no interaction between language users, feedback, and effect of language use" (Bachman). Relationships between the input and the response are adaptive "If the input is influenced by the response, but without the feedback that characterizes a reciprocal relationship" (Bachman, p. 151). This means, for example, a test taker's result in one specific task or a test item will determine his/her next tasks and items that fit the test taker's level (Bachman, p. 154).

Bachman (1990) also stated that test method facets should be systematic, which means tests are consistent in terms of forms. For example, if the test is designed in multiple-choice format, it should be multiple-choice, not a different format, when it is given another time. Test constructors

should be aware of influences of these test methods facets and minimize their impact for the sake of testees' best performance.

The next factor that communicative test makers should pay attention to is attributes of individuals. Attributes of individuals involve not only individual characteristics but also group characteristics. Individual characteristics usually concern cognitive style and knowledge of specific content areas, while group characteristics may include sex, race, and ethnic background. Like Danili and Reid (2006), Bachman (1990) was concerned about the effects of the testee's cognitive style on their test performance. He pointed out that cognitive styles such as 'field dependent/field independent' and 'convergent/divergent', correlate to and affect test takers' performance to some degree. An example of knowledge of education is that "knowledge of economics is likely to affect an individual's performance on any test in which economics is included as propositional content" (Bachman, p. 164). In fact, personal attributes also constitute a source of error in the measurement of communicative language ability.

Random, or unsystematic, factors also affect an individual's test scores. Random factors refer to unpredictable and largely temporary conditions, for example, testees have varying degree of mental alertness, or uncontrolled differences in test method facets (e.g., the test is postponed to another day), or idiosyncratic differences among test administrators when they give the test. Some other unsystematic factors can be the imprecision of the scales, incompleteness of language sample, or limitations on observation and quantification. Indeed, these examples of random factors warn test makers that interpretations of test takers' performance based on test scores may not provide an accurate measurement of their language ability.

In short, some of the factors that potentially affect examinees' performance might be test method facets, attributes of the test taker, and random factors. Whether these factors are labeled systematic or unsystematic factors, they are considered sources of measurement error. Therefore, testers should be aware of their impact, and they should know that making inferences about one's language level ability based on his/her test score is not always accurate and reliable. The more the extraneous factors are minimized, the higher the accuracy of assessment and reliable inferences by test makers are made.

Bachman's (1990) and Danili and Reid's (2006) lists of factors that affect test takers' scores are convincing and helpful for educators, teachers, and test makers who have interest in or concern about language tests in general and communicative language tests in particular. However, I could not find in the literature solutions to minimize the influence of these factors. In my opinion, test makers will have to cope with many difficulties if they want to reduce the effects of elements other than language communicative ability in test performance because some factors are in the test writers' control but others are not. This issue deserves further research to help test takers take the most advantage of their ability when taking the test.

Challenges in Communicative Testing
As a matter of fact, communicative testing is a challenge for test designers. One reason is the issue of predictive validity. When designing a test of communicative ability, identifying test takers' needs based on communicative encounters that they are likely to experience is one of the basic principles. However, it is not certain if test makers can guarantee that testees performing well on a test in class are also able to do well outside the classroom in a real life situation. One reason for this is that real life communication is characterized by unpredictability. Studies have proved that test designers have tried to make real-world tasks, but encountered difficulties from the varied or diverse nature of contexts (Katsumasa, 1997; Brown, 2003).

Conclusion
In summary, I have discussed four principles of communicative test design based on a communicative view of language competence built by a team at the OISE. The

principles are start from somewhere, concentrate on content, bias for best, and working on washback. A communicative test offers communication meaningful for learners in real-world contexts where students experience and produce language creatively using all four language skills of listening, speaking, reading, and writing. It is important to note that whether test scores are reliable or not really depends not only on extraneous factors aside from test takers' language ability but also on raters' training with relevant scoring criteria and procedures.

Implementation of communicative tests is feasible and it promises positive effects on English training and teaching in EFL contexts such as Vietnam. In this particular country, communicative testing benefits from the presence of the communicative teaching approach which is already in place in most college-level English programs. In some teacher-training schools such as Can Tho University, the communicative language teaching approach is quite successful, and it provides students better command of English as well as teaching methods that contribute to their excellent reputation for achievement in their teaching when they graduate (Nguyen, 2000). It is

reasonable and necessary to implement communicative tests in harmony with the communicative teaching method to assess learners' ability to communicate. However, according to Nguyen and Vu (1999) who performed a study on the application of communicative test at the University of Economics, Vietnam, in order to use the communicative test effectively, the staff really needs some training to have the necessary language testing expertise. They also recommended that there needed to be more studies of test elements, such as test tasks, focus, and content because these elements contribute to the validity of a test. In my view, it is good to implement communicative testing in Vietnam because it helps teachers measure learners' language ability more accurately. It also helps learners become familiar with the kind of testing that they may encounter when they take international tests, such as the TOEFL iBT or the IELTS. Using communicative tests in Vietnam is an important and necessary change from traditional tests, which are grammar-based. I strongly believe that communicative test constructors in Vietnam will be able to find ways to implement this new kind of test successfully once they are aware of its limitations as mentioned above.

References

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

Bailey, K. M. (1998). *Learning about language assessment: Dilemmas, decisions, and directions*. Pacific Grove, CA: Heinle & Heinle.

Brown, H. D. (2003). *Language assessment: Principles and classroom practices*. White Plains, NY: Longman.

Brown, J. D. (2005). *Testing in language program*. New York: McGraw-Hill.

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

Carroll, B. J. (1983). Communicative language tests: Tasks, enabling skills, formats, and measurement criteria. *World Language English, 2*(1), 37-39.

Danili, E., & Reid, N. (2006). Cognitive factors that can potentially affect pupils' test performance. *Chemistry Education Research and Practice, 7* (2), 64-83. Retrieved December 4, 2007, from http://www.rsc.org/images/DaniliReid%20final_tcm18-52108.pdf

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

Katsumasa, S. (1997). Communicative language testing: Principles and problems. *English Review, 12*, 3-24. Retrieved December 1, 2007, from http://ci.nii.ac.jp/naid/110004 693443/en/

Nguyen Thi Thanh Ha & Vu Thi Phuong Anh. (1999). *Is there communicative language testing at the University of Economics?* Paper presented at the Fourth Interna-

tional Conference on Language and Development, Hanoi, Vietnam.

Nguyen Thu Huong. (1999). *Some reflections on the introduction and application of the communicative approach in the process of developing the English language in the Mekong Delta.* Paper presented at the Fourth International Conference on Language and Development, Hanoi, Vietnam.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second lan-*

guage performance assessments. Honolulu: Second Language Teaching & Curriculum Center, University of Hawaiʻi at Manoa.

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

Wesche, M. B. (1983). Communicative testing in a second language. *The Modern Language Journal, 67* (1), 41-55.