Testing Communicative Language Skills:
An Investigation of Practices and Assumptions

コミュニケーション能力テストにおける実践調査

Roger Palmer

大学生の外国語におけるコミュニケーション能力は、常にテストによって答えが出るとは限らない。しかし、教師は教えた事が生徒にどの程度理解されているかテストをするべきではある。テストを与える事により学習の動機を与え、学習者をサポートし、言葉により何かできると言う自信を与える事もできるであろう。そしてまた、全てのクラスで同じテストが同じように効果的であるとも言えない。それを、関西大学の学生に与えた一連のテストと結果を基に、説明・論説したものがこの研究のメインテーマである。これが、これからテストや課題を作られる先生方の役に立つことを願っている。

Introduction

This paper investigates ways that communicative language skills are tested, beginning with the types of tests most commonly discussed in current research and of relevance to university education. It then seeks to clarify the assumptions that typically underlie the selection of appropriate test models, in an effort to discover those that are most suitable - in terms of being effective, reliable and valid - for language testing. Examples of tests administered to students in first-year English communication classes at Kansai University are discussed and subjected to statistical analysis to determine whether their results satisfy the criteria they claim to test. The paper concludes by offering tentative conclusions about good testing practices based on current research findings, yielding results that can be applied to aid the teaching of communicative language.

i Testing practices, and their relevance to the testing of communicative language

Reading through the research on testing can suggest test types that match the needs of the class, but potential pitfalls abound. While neat labels for test types exist, any one test will normally combine the characteristics of more than a single category. Moreover, there is no complete objectivity in testing. The frequently administered multiple-choice test is a case in point: even

when objectively marked, decisions on the setting of questions, writing of correct choices, and inclusion of other distractor answers are subjective. Rather than worrying about good or bad tests per se, frequently it is more pertinent to consider the specific context to which the test is applied (Underhill, 1987, p.6).

For oral communication classes, **Direct** tests involving spoken samples of actual language performance can be considered the most accurate for testing proficiency in speaking and listening. Interviewing the students is real and uncontrived, but is nevertheless time-consuming and costly. **Indirect** testing is thus more common, with communicative ability expressed on paper. Such tests boast utility value, allowing for measurement statistically in terms of reliability or predictive validity (Henning, 1987, p.5). By *reliability*, what is meant are errors of test method (the location, unclear instructions, time pressure, etc.) as well as errors independent of the ability which is supposed to be measured (due to tiredness, motivation, test strategy, etc.). By *validity*, what is meant are errors of interpreting and using test scores, as well as the necessity of showing that a test measures what it says it measures.

Many diagnostic tests are **discrete-point**, that is, they test performance in a narrow area of the language (such as a focus on preposition use only in a cloze); whereas **integrative** tests (a random cloze or dictation, for example) are designed to measure a variety of language abilities at the same time, and point to overall proficiency (Henning, 1987, p.5). **Aptitude** tests can determine student suitability for a particular program. Vocabulary is a good indicator of aptitude, but a poor element for testing: for it lacks face validity in comprehension, and has a bad 'backwash' effect on classroom practice. Vocabulary tests reveal intelligence or knowledge of the topic, not communicative ability or benefit from instruction. Likewise, **proficiency** tests (TOEIC, etc.) tend not to be drawn from the teaching on a particular course, and are best used for context-specific placement and selection (Henning, 1987, p.6). Streaming within a university based on proficiency tests purports to be a tried-and-tested formula, but tends to be largely irrelevant to the teaching students have received in any one institution (i.e. in the target-language domain), and hence should be treated as one of the most controversial practices operating at present.

Achievement tests are for program evaluation, directly drawn from the content of instruction, and show if students have learnt what has been taught (Henning, 1987, p.6). Class tests are often **criterion**-or **domain-referenced**. Criteria for each level of achievement are stated, with tests related to teaching objectives, hence teachers are likely to want their students to obtain a

high score; but with no norm to compare to, bright students might not improve. By contrast, **norm-referenced** or **standardised** tests are easy to compare, reliable, valid and replicable; but being independent of instruction, they fail to match the taught course objectives.

ii Assumptions that underlie testing

Current theories no longer assume that language is primarily about structures (e.g. at the level of syntax) requiring tests of isolated language components; on the contrary, language tends to be viewed as a way of carrying out functions or communicating meanings, therefore relevant test measurements show what learners can do with language (Hughes, 2003, Ch. 10). Tests are driven by the teaching that has occurred, supporting the learning process and motivating students. Tests operating under the old assumptions, constructed mainly for assessing students' performance in the L2, result in teaching driven by the test (Heaton, 1988, p.5). Focusing too greatly on testing language elements tends to have a detrimental effect on communicative teaching of the language. Fluency, or the ability to communicate in a range of situations closely related to real life, shows that the student can use the language when speaking and listening, whereas a test of whether a student can manipulate certain structures effectively does not mean they are proficient speakers (Heaton, 1988, p.10). Assessing language skills instead of structures demands sampling, for while there is not enough time to test everything that has been taught, the test must fairly represent the kinds of skills that the students learnt (Heaton, 1988, p.12). A language test, therefore, must not trick them into giving wrong answers. Indeed, the inclusion of language elements that have not been taught may trap the more able students (Heaton, 1988, p.14). Recognition tests, such as multiple choice, are especially prone to traps and hence misleading results, whereas production tests, in which students are asked to complete sentences, for example, have a number of possible correct answers which demonstrate the extent of a learner's proficiency. Integrative tests, such as a cloze, benefit from requiring students to use background, linguistic and textual knowledge.

Testing communicative ability has given rise to the use of bands with descriptions for each. Speaking tests measure the authentic use of language for communication, and are hence highly valid. However, reliability cannot be assumed as examiners are likely to get tired. Furthermore, the scoring of speaking tests tends to be made on the spot (unless even more time is spent reviewing tapes and videos of tests), and will be subjective and variable, again affecting reliability. Speaking by its nature being interactive, examiners will react differently to individuals, and any interview in fact is a speaking *and listening* test (Hughes, 2003, Ch.10).

The next section will describe tests that were administered to approximately 600 first-year students taking the English Communication 1 course at Kansai University. Questions were designed to reflect the kinds of study that had been undertaken in class. Test items had to discriminate between different language abilities within each class of 30 individuals, as well as identifying a range of communicative abilities across classes studying similar content, including those majoring in commerce, economics, engineering, letters, law and sociology. It was decided that a comprehensive and representative sample of the communicative language skills of these learners should include the four broad skill areas of speaking, listening, reading and writing, which are treated in turn below.

iii Tests administered at Kansai University

Speaking (and listening)

Examinees took part in both simple dialogues and multi-participant interactions, designed to let learners reveal the extent of their ability to comprehend and produce language (Hughes, 2003, Ch.10). They were observed and placed into speaking bands.

During a 'Find Someone Who' task, in which students asked follow-up questions for extra information, individuals were rated as they interacted in pairs. Then they were then rated in groups of four as they took part in a vocabulary card game. Finally, they were given an information gap task. All language in the tasks was recycled from previous class instruction, being geared towards demonstrating the extent of their skill in sustaining conversations in the target language.

Speaking ratings

Student number	Band
1	1
2	1-
3	1
4	1
5	1
6	1-
7	1
8	1

9	1
10	1

(Results are given for the first 10 students on the class roll.)

According to Weir (1990, in Mangubhai, 2004, p.5.47), a band of '1-' represents:

accuracy	pronunciation unintelligible
appropriacy	unable to function
range	unable to express meaning
flexibility	one-word response or no response
size	no utterance at times

whereas a band of '1' signifies:

accuracy	pronunciation heavily influenced by L1 but intelligible
appropriacy	broadly able to function
range	severely limited
flexibility	unable to initiate conversation
size	one or two simple utterances.

Oral bands have the advantage of making assessment criteria explicit, reducing the degree of subjectivity and enhancing the level of reliability. Validity is less of a problem, since these oral tests are a measurement of the communicative abilities they set out to test. Note the limitations of the testing process here: speaking and listening are not isolated from each other; and results reflect abilities that may have been acquired before the course started, as well as during the course.

This mid-term speaking test reveals the low proficiency of these learners (first-year non-English majors in engineering), which was largely as expected. Of greater significance is that the test results are skewed by the limitations of the oral language proficiency ratings themselves. A subsequent test produced more specific data in its rating scales, by using the ACTFL Proficiency Guidelines - Speaking (Hadley, 2001, pp. 471 - 6). The oral test used above identified those students at level 1- and level 1, but it proved too imprecise for the purposes of matching classroom content to the needs of the learners. The ACTFL, on the other hand, subdivides learners described here as 'Band 1' into more precise descriptors (Novice High, Intermediate Low, etc.), creating a good backwash effect that instructors can apply in the classroom.

Listening

Dictation is a useful measurement of listening, signalling familiarity with the grammatical and lexical patterning of the language, and overall textual comprehension (Heaton, 1988, p. 17). It can be described as an integrative test in that it deals with a range of linguistic challenges in a single task. The partial dictation format is easier to mark reliably than a traditional dictation (Hughes, 2003, p. 168). It is an objectively marked test, seeking to limit the confusion over ambiguous answers that can plague multiple choice exercises.

In partial dictation scoring, listening skills can be separated from the ability to spell correctly, although it is vital to mark as correct only those answers that demonstrate the learner has made sense of the sounds they heard. The difficulty in marking lies in ascertaining whether the student did indeed hear and recognise the word in the cases when they misspelt it.

Scoring key:	Wrong choice of word e.g. natural reader; is infuse-iastic	deduct a point
	Spelling mistake e.g. is inthusiastic, is enfusiastic	correct

Listening Test: Teacher's Script with Answers in Bold Type

There are 16 types of personality. Type 1 is serious, quiet, and wants a peaceful life. Type 2 is reserved and **(1) interested (2) in** how things work. Type 3 is kind, hardworking, and dependable. Type 4 **(3) is (4) sensitive** and does not like conflict. Type 5 is quiet but forceful and original. Type 6 is idealistic and interested in helping people. Type 7 is independent, original, and a **(5) natural (6) leader**. Type 8 is logical and creative, but hard to get **(7) to (8) know** well. Type 9 is friendly, adaptable, and an active person who looks for quick **(9) quick (10) results**. Type 10 is practical, traditional, and often athletic. Type 11 loves people and fun, and has common sense. Type 12 is **(11) warm- (12) hearted**, popular, and hardworking, Type 13 **(13) is (14) enthusiastic**, idealistic, and creative. Type 14 is popular and sensitive, and dislikes **(15) being (16) alone**. Type 15 is creative, resourceful, and enjoys **(17) enjoys (18) friendship**. Type 16 is a leader, good at public speaking. Which **(19) do (20) you** think you are? (160 words)

The difficulty level of the test/item was determined using facility values (***p*-values**):

to find the difficulty level of item X,

count the number of students who got X correct and divide that by the the total number of students who sat the test.

The answer (p-value) is the proportion of correct answers.

Table 1: Listening

S No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	score	
1	1	0	0	1	0	1	0	0	1	1	0	1	0	0	0	1	1	1	1	1	11	
2	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	7	
3	0	1	1	1	1	1	0	0	1	1	0	0	1	0	1	1	1	1	0	0	12	
4	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	17	
5	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	16	
6	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	1	1	1	1	1	10	
7	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1	1	13	
8	1	1	0	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	1	16	
9	1	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0	1	1	1	1	11	
10	1	1	1	1	1	1	0	0	0	0	1	0	0	0	1	1	1	1	0	0	11	
11	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	1	1	1	0	0	12	
12	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	7	
13	1	1	0	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1	0	0	14	
14	0	0	1	1	0	1	1	1	0	0	0	0	1	0	0	0	1	1	0	0	8	
15	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1	1	1	0	0	6	
16	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
17	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	0	1	1	1	1	13	
18	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	15	
19	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	8	
20	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	10	
21	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	16	
22	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	18	
23	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	7	
24	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	11	
25	1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	1	1	1	0	0	13	
26	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	10	
27	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	9	
28	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	14	
29	0	1	0	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	0	13	
sum correct	19	20	21	28	17	24	13	15	12	12	16	8	11	1	11	18	27	27	16	14	11	
p	0.7	0.7	0.7	1.0	0.6	0.8	0.5	0.5	0.4	0.4	0.6	0.3	0.4	0.0	0.4	0.6	0.9	0.9	0.6	0.5		
L gp	4	5	7	9	3	6	4	5	2	2	1	0	1	0	0	2	8	8	4	3		74
H gp	9	8	7	10	6	10	8	8	7	7	10	5	6	1	6	8	10	10	8	8		152
DI	0.5	0.3	0	0.1	0.3	0.4	0.4	0.3	0.5	0.5	0.9	0.5	0.5	0.1	0.6	0.6	0.2	0.2	0.4	0.5		

The listening test for a class of 29 students had an average score of 57% (11.4 out of 20), with the highest being 90% (18 out of 20) and the lowest 10% (2 out of 20). Q1 was easy, Q7 and Q8 mid-range; and Q14 much too difficult. Item 10 (*results*) was frequently mistaken for *resorts*. Students should have been familiar with the topic of personality types and lexical items, as they had been introduced during the course.

Item discriminability is defined as the ability of a test to discriminate between weaker and stronger examinees (Henning, 1987, p. 51). If discriminability ranges from zero to one, then it is

important for the test designer to decide where acceptable discriminability begins, e.g. at one-third or two-thirds of the way along a line which represents the discriminability continuum (ibid., p. 52).

Discrimination Index (Henning, 1987, p. 52)

DI = No. correct in H - No. correct in L

No. of students in H

(H = the 10 students with highest total scores on the test;
L = the 10 students with lowest total scores on the test)

Q3 would normally be rejected on the grounds of parity in L group and H group answers; likewise Q4 (L9, H10) and Q14 (L0, H1) suffer from poor discriminability for quite different reasons: Q4 is too easy and Q14 too difficult. Also, Q17/18 (L8, H10) are too poorly differentiated along the discriminability continuum. With those 5 exceptions (25%), the remaining 75% of the questions exhibit a noticeably higher discriminability on the whole than for the cloze reading (see below). Caution is urged at leaping to hasty conclusions, for at low proficiency levels there may be a general inability to cope with the demands of production in a cloze test, whereas slight differences in listening ability or lexical knowledge may be magnified in a dictation.

Reading

In the communicative classroom, learners are required to read, comprehend and apply classroom instructions in situations where they need to clarify meaning. Cloze passages test reading in ways that are indicative of overall ability (Hughes, 2003, p. 193), and hence form another example of integrative testing. The test that was administered below presented a number of challenges that had to be taken into account: matching the difficulty level of the passages to the learners through trial and error; relating the conversational style and examples of classroom language to class content; including a longer passage of uninterrupted text followed by gaps, at random or targeted; predicting ways of filling in the gaps; writing and revising instructions for clarity; adjusting the layout to enhance ease of marking; giving students some experience of the cloze format beforehand; and validating results in relation to listening, writing and speaking band scores (Hughes, 2003, Ch.10).

Being primarily a test of communicative reading ability, the content avoids culturally-orientated texts requiring background knowledge. Any grammatically correct English that makes sense in the context is acceptable, whether spelt correctly or incorrectly. One point is awarded for each correct answer, irrespective of whether it is one word or more than one word.

Reading Test

Part 1. Read the 5 short conversations. In each space numbered from 1 to 10, write a word (or words) that completes the sentence. *Any words are acceptable if the sentence makes sense in English.*

Example: Where _____ you live?

Acceptable answers: do; will; would, etc.

Conversation 1

A Do you know what **1** _____ the class starts?

B I think it begins at **2** _____.

Conversation 2

A Have you done the homework?

B No, I was too **3** _____.

Conversation 3

A What should I do when I don't understand?

B You should just say, 'How do you **4** _____ **5** in English?

Conversation 4

A What must I do when I'm late for class?

B You have to say, **6** _____ **7** I'm late.

Conversation 5

A What's your personality **8** _____?

B Oh, I'm kind and **9** _____.

A How **10** _____ you?

Part 2. Read the passage. In each space numbered from 11 to 20, write a word (or words) that completes the sentence. *Any words are acceptable if the sentence makes sense in English.*

Many students ask what the best way to learn English is. One good method is to talk a lot to English speakers. **11** _____ in Japan this is not easy because almost **12** _____ students speak to Japanese speakers. One suggestion is to **13** _____ to a pen pal. Although it is not the same as speaking, communicating by mail or e-mail can be really **14** _____. To practise listening skills, it is a good idea to **15** _____ movies and **16** _____ to English pop music. Try to find the words to **17** _____ songs if you can. Reading is **18** _____ excellent way to learn. In fact, reading will **19** _____ your general English level to improve. Good **20** _____ in your studies!

Table 2: Reading

S No.	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	score	
1	0	0	1	0	0	1	1	0	1	1	1	0	0	0	1	1	1	0	0	0	9	
2	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	0	0	1	16	
3	0	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1	0	1	1	14	
4	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	14	
5	1	1	0	1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	0	1	14	
6	1	1	0	1	0	1	1	0	1	1	1	0	0	1	1	1	1	0	0	1	13	
7	1	1	1	0	0	1	1	0	1	1	1	0	0	1	1	1	1	0	0	1	13	
8	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	0	0	1	12	
9	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	15	
10	0	0	1	0	0	1	1	0	1	1	1	1	0	0	1	1	0	0	0	0	9	
11	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	10	
12	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	1	0	0	0	11	
13	1	0	0	0	0	1	1	0	1	1	1	0	0	1	1	1	0	0	0	1	10	
14	0	1	1	1	1	1	1	0	1	0	1	1	0	0	1	1	1	0	0	1	13	
15	0	1	1	0	0	1	1	1	0	1	1	0	1	0	1	1	0	0	0	0	10	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	14	
18	1	1	1	0	1	1	1	1	0	0	1	1	1	0	1	1	0	0	0	1	13	
19	1	0	1	0	1	1	1	0	1	1	1	0	0	0	1	1	0	0	0	0	10	
20	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	14	
21	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	17	
22	1	1	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	0	0	14	
23	0	1	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	7	
24	1	0	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	0	0	0	13	
25	1	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	7	
26	1	1	0	1	1	1	1	0	1	1	1	0	0	0	1	1	0	0	0	1	12	
27	1	1	1	1	0	1	1	0	1	0	1	0	0	0	0	1	1	0	0	1	11	
28	1	1	1	0	0	1	1	0	1	1	1	1	0	0	1	1	1	0	0	0	12	
29	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0	1	1	14	
sum	20	19	22	15	15	27	26	11	22	23	24	12	8	12	24	24	18	2	3	14	12	
p	0.7	0.7	0.8	0.5	0.5	0.9	0.9	0.4	0.8	0.8	0.8	0.4	0.3	0.4	0.8	0.8	0.6	0.1	0.1	0.5		
Lgp	5	5	7	3	2	9	9	3	6	8	6	1	1	1	5	6	2	0	1	3		83
Hgp	8	8	8	8	9	10	8	6	8	9	10	6	4	6	10	10	9	2	2	5		146
DI	0.3	0.3	0.1	0.5	0.7	0.1	-0.1	0.3	0.2	0.1	0.4	0.5	0.3	0.5	0.5	0.4	0.7	0.2	0.1	0.2		

The reading test was subjected to data analysis using facility values to check the level of difficulty of items, and a discrimination index to identify the reading ability of each examinee.
The test had an average score of 59% (11.8 out of 20), with the highest being 85% (17 out of 20) and the lowest zero. Normally a p value above 0.67 is too easy and below 0.33 is too difficult. On that basis, Q1 was too easy; Q4/5 mid-range; and Q18 and Q19 much too difficult. The test followed the normal practice by starting off with easy questions to virtually ensure that even the weakest students could score some marks, with a few much harder questions to discriminate amongst the better students.

By these measures, the reading test suffers from too little discriminability for 8 items or 40%. At first sight, it might make sense to discard Q3, 6, 7, 9, 10, 18, 19, and 20, leaving the remaining 12 items (60%) out of the original 20. However, several items that suffered from poor discriminability had already been taught on the course. The ability to isolate potentially difficult items means the instructor may be able to pre-teach them more thoroughly another time. In deciding which items to remove from the test, teachers are forewarned not to leap to hasty conclusions by misinterpreting or misapplying data. That being said, evaluating students without evaluating the tests themselves is the critical issue here, since tests need to respond to student needs.

Writing

A writing test has to take account of a representative sample of tasks so that a range of writing uses is covered. The example given below was designed for an oral communication class. Students were asked to write informally without dictionaries (10 minutes' freewriting) on the topic. Informal writing resembles the students' efforts at spoken communication, albeit with the chance to briefly reflect on - though not time to review and significantly improve - the accuracy of sentences which would otherwise sound disjointed in spoken discourse. Together this forms an integrative approach to testing language in context, with a primary emphasis on meaning and the overall effect of discourse on communication (Heaton, 1988, p.16).

The scoring system, though subjective, attempts to be transparent and consistent.

Content quality and quantity	4
Organisation	3
Language use including vocabulary	2

Mechanics including grammar　　1

The scoring reflects the communicative effort, not the level of control over basic structures of the language.

Writing Test

Instructions

Write for 10 minutes on the following topic:

What kind of person are you?

Conclusion

The testing of communicative language skills has implications for raising the standards of language instruction. As current theories of language have moved beyond a definition of language as simply a set of structures, so testing needs to motivate students and support the learning process. One solution posited here has been to adopt a kind of 'integrative' approach to testing, that is, one that seeks to measure a variety of language abilities in a time- and cost-efficient manner. Without a doubt, data produced by tests will normally be reflected in student evaluation, but grading as a sole (or, for that matter, main) justification for testing is not supported by the research. Subjecting test results to statistical analysis is recommended to ensure that validity and reliability are placed at the heart of test construction. It is envisaged that evaluating the practices and assumptions of the tests administered will serve the student population more fairly. Reservations have been expressed about some common testing practices and assumptions, notably an overdependence on multiple choice testing and its objectivity; the use of proficiency scores to measure students within an institution when those scores do not reflect the kinds of instruction that students have received; and the conducting of speaking tests without measurable bands, or with inappropriate rating scales, or by examiners lacking adequate knowledge of those bands. Indeed, any test applied out of context is prone to defeat the objectives for which it was intended; and putting the learners at the heart of language acquisition and language testing promises to be a much more effective starting-point. By challenging assumptions that hold back effective testing, it seems reasonable to assume that communicative tests will be a positive addition to the teacher's repertoire, and will help to improve students' communicative skills.

Bibliography

Bachman, L.F. (1990). Fundamental considerations in language testing. Oxford: OUP

Hadley, A.O. (2001). Teaching language in context. Boston: Heinle & Heinle

Heaton, J.B. (1988). Writing English language tests (new ed.). London: Dover

Henning, G. (1987). A guide to language testing: development, evaluation, research. Boston: Heinle & Heinle

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

Underhill, N. (1987). Testing spoken language. Cambridge: CUP

Weir, C.J. (1990). *Communicative language testing*. London: Prentice Hall International, quoted in Mangubhai, F. (2004). Language Testing. Queeensland, Australia: USQ