ABSTRACT
        A theoretical framework that distinguishes the
knowledge, or competence, aspect of language proficiency from the
skills aspect is outlined, and the factors in the language testing
situation that affect performance on language tests are examined. The
model is intended for use in assessing the construct validity of
tests of language proficiency and in clarifying some terms and
concepts that have been used to describe various aspects of language
proficiency. The nature of the performance tasks and language
competencies required by some widely used testing procedures are then
examined within this framework, and the implications for psychometric
theory and language testing research are discussed. (MSE)

AN EXAMINATION OF SOME LANGUAGE PROFICIENCY TESTS FROM A COMMUNICATIVE
VIEWPOINT

Lyle F. Bachman (University of Illinois at Urbana-Champaign)

1   Introduction

    The theme of this conference, "Language = Knowledge and Skill,"
is particularly well-chosen, since it represents, I believe, a
recognition that language proficiency involves both knowledge, or
competence, and skill in implementing, or executing that competence.
Skills and components models such as those proposed by Lado (1961) and
Carroll (1961) distinguished skills (listening, speaking, reading, and
writing) from components of knowledge (grammar, vocabulary,
phonology/graphology), but did not indicate how these were related.
It was not clear whether the skills were simply manifestations of the
knowledge components in different modalities and channels, or whether
they were qualitatively different in some other ways. For example,
does reading differ from writing only in that it involves reception
rather than production? If that were so, how can we account for the
fact that quite competent and skillful readers are not always skillful
writers? Chomsky's model (1965), with its distinction between compe-
tence and performance, permitted us to distinguish random "noise" from
language proficiency, but in so doing limited language proficiency
solely to competence. And neither of these models recognized the full
context of language use--the contexts of discourse and situation.
Halliday's framework (1976), with its focus on functions, both illocu-
tionary and textual, clearly recognizes the context of discourse, but
again is limited to competence. Finally, although Hymes' (1972)
notion of sociolinguistic appropriateness recognizes the interaction
between language use and the context of situation, it does not address
the distinction between competence and skill.

    Recent frameworks of communicative competence (Munby, 1978;
Widdowson, 1978; Canale & Swain, 1980; Savignon, 1983), provide a
much more inclusive description of the the knowledge required to use
language, in that they incorporate linguistic competence, discourse
competence, and sociolinguistic competence. All of these frameworks
comprise what might be called descriptive rather than working models
in that they focus on competence and either explicitly or implicitly
ignore the implementation of that competence in language use. A more
cognitive approach to language use has been taken in working models of
language processing such as those proposed by Faerch and Kasper (1983)
and Bialystok and Ryan (forthcoming). But while such models
distinguish planning from execution and characterize varying degrees
of cognitive control in language processing, they do not specify how
language competencies are distinguished from language skills.

    At this point it may be useful to discuss what evidence there is
for a distinction between competence and skill. The first kind of
evidence lies in the differential implementation of competence in
different, skill modes. It is well-known, for example, that
comprehension generally precedes production, both in first and in
second language acquisition. If the requisite competencies are

2

present for comprehension, what causes this implementation in
production to lag behind? Likewise with differences in channel.
Fluent reading does not necessarily imply fluent listening, or vice-
versa. Another kind of evidence can be found in the literature on
communication strategies, which deals with the ways in which
individuals attempt to achieve some communication goal in a given
language despite inadequate communicative competence in that language.
In such cases the individual may be able to compensate for inadequate
competence by the implementation of various strategies. Thus, the
most skillful communicator may not necessarily be the individual with
the greatest competence.

To assure you that I am not merely juggling terms, I will
a theoretical framework that distinguishes the knowled
competence aspects of language proficiency from the skills asp
and that also addresses the factors in the language testing situ ...m
that affect performance on language tests. I believe this model is
of use in examining the construct validity of tests of language
proficiency and in clarifying some terms and concepts that have been
used to describe various aspects of language proficiency. I will then
examine, within this framework, the nature of the performance tasks
and language competencies required by some widely used testing
procedures. Finally, I will discuss the implications that this
examination suggests for psychometric theory and for language testing
research.

2 A framework for describing performance on tests of language proficiency

Adrian Palmer and I have proposed a framework for describing the
different factors that affect performance on language tests (Bachman
and Palmer, forthcoming). This framework includes four types of
factors: language trait factors, skill factors, method factors, and
random factors. Language trait factors are those competencies or
mental abilities that are specific to language use, and are of two
main types: organizational competence and pragmatic competence.
Organizational competence, which includes grammatical and discourse
competence, pertains to the formal characteristics of language usage.
Pragmatic competence, which includes illocutionary and sociolinguistic
competence, pertains to the functional and social characteristics of
language use.

Skill factors are those general characteristics of the individual
that affect test performance. These consist of 1) psycho-
physiological skills, which are distinguished in terms of mode
(productive/receptive) and channel (aural-oral/visual), 2) forms of
representation (conscious/subconscious, analyzed/unanalyzed, pre-
fabricated routines, rules), which determine the extent to which
language competencies are available for use, and 3) strategic
competence, which consists of a set of general abilities that affect
how language competencies are implemented for maximum effectiveness in
processing information.

Method factors are those characteristics of the test method that effect performance. These factors consist of 1) the type of language use situation (reciprocal/nonreciprocal), 2) the amount of context (embedded/reduced), 3) the distribution of information (compact/diffuse), 4) the type of information presented (concrete/abstract), and 5) the type and degree of restrictions on language performance; these include restrictions on the organization of discourse, the language use situation, propositional content, illocutionary force, forms, participants, mode, channel and time/length. "Communicative" testing methods might be characterized as those involving relatively unrestricted, appropriately contextualized language performance, while "non-communicative" testing methods involve only artificially restricted, inappropriately contextualized language performance.

Finally, random factors consist of 1) cognitive and affective characteristics of the individual, such as field dependence/independence, inhibition, tolerance/intolerance of ambiguity, and motivation, 2) interactions among specific combinations of trait, skill and method factors, and 3) measurement error.

One application of this framework has been in the definition of terms that have been used to refer to various aspects of language proficiency (Bachman and Palmer, 1984). Linguistic competence __, be defined as consisting of the trait factors of grammatical competence: syntax, morphology, and phonology/graphology, while communicative competence consists of linguistic competence plus the trait factors of discourse competence, illocutionary competence, and sociolinguistic competence.
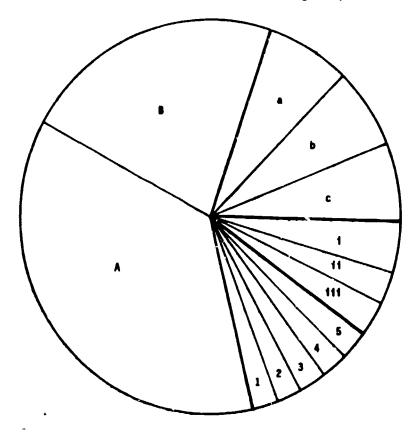
Language skills (listening, speaking, reading, writing) consist of trait and skill factors. Linguistic performance consists of the manifestation of linguistic competence and skill factors in artificially restricted and inappropriately contextualized test situations (methods). Communicative performance consists of the manifestation of communicative competence and skill factors in relatively unrestricted and appropriately contextualized test situations (methods).

A measure of linguistic performance includes that portion of a test score attributable to linguistic competence, skill factors, artificially restricted and inappropriately contextualized method factors, and random factors. A measure of communicative performance includes that portion of a test score attributable to communicative competence, skill factors, relatively unrestricted and appropriately contextualized method factors, and random factors.

This framework may also be useful in explaining sources of variation in performance on tests, as illustrated in the Figure below.

I. Trait Factors:
  A. Organizational Competence
  B. Pragmatic Competence

II. Skill Factors:
  a. Psychophysiological
  b. Forms of Representation
  c. Strategic Competence

III. Method Factors:

1. Language Use Situation
2. Amount of Content
3. Distribution of Information
4. Type of Information
5. Restrictions on Language Performance

IV. Random Factors:

i. Cognitive and Affective Qualities
ii. Interactions among other Factors
iii. Measurement Error

Sources of Variation in Language Test Scores

5

The relative contribution of trait, skill, method and random factors
to test performance will, of course, vary from test to test and from
individual to individual. For example, Bachman and Palmer (1982)
found that a multiple-choice test of grammatical competence loaded
much more heavily on the method factor than did multiple-choice tests
of either pragmatic or sociolinguistic competence. The effect of the
task on test performance has generally been dealt with
psychometrically as systematic error variance associated with the test
method (Campbell & Fiske, 1959). The framework described here
specifies in more detail the factors that comprise test method and
at the same time recognizes the relationship between the demands set
by the task and context of the test and the competencies required to
successfully meet these demands.

Finally, this framework may be useful in clarifying some
misconceptions regarding the terms "direct" and "indirect" as they
have been applied to language tests. The term "direct test" is often
used to refer to a test method in which performance resembles "actual"
or "normal" language performance, while an "indirect test" is one in
which test performance is perceived as somehow different from "actual"
or "normal" performance. Thus, writing samples and oral interviews
are referred to as "direct" tests, since they presumably involve the
use of the skills being tested. By extension, such tests are often
regarded, virtually without question, as construct valid and therefore
as legitimate criteria for the validation of "indirect" tests.

There are two problems with this, however. First, we have no
definition of "actual" or "normal" language use that is precise enough
for us to determine the extent to which performance on a given test is
similar to such language use. Indeed, the framework suggested here
may at best permit us only to distinguish relatively "communicative"
from relatively "non-communicative" language performance. A more
serious problem, however, is that the use of the term "direct"
confuses the behavioral manifestation of a trait or competence for the
construct itself. As with all mental measures, language tests are
indirect indicators of the underlying traits in which we are
interested. The framework presented above captures this distinction
by recognizing that there are factors in addition to trait factors
that affect performance on all language tests, whether these require
recognition of the correct alternative in a multiple-choice format or
the writing of an essay.

3   An examination of some tests of language proficiency

In examining proficiency tests as measures of communicative
performance, there are two questions that should be addressed.
First, to what extent do the tasks required on the test involve
communicative language performance? Second, to what extent does the
test assess communicative competencies?

3.1   Multiple-choice tests

The multiple-choice test is one of the the most widely-used
techniques for testing language proficiency in the world. Such tests
typically include parts aimed at measuring at least some of the

following skills or components: 1) listening comprehension, 2)
structure, and 3) reading comprehension. The language performance
tasks on such tests are almost always restricted to non-reciprocal
situations, in which there is no potential for feed-back or
negotiation of meaning. The amount of context generally varies
considerably from part to part, as does the distribution and type of
information. The format of these tests generally restrict the mode of
performance to reception. Finally, there are obvious restrictions on
time and length.

3.1.1 Multiple-choice tests of listening comprehension

Multiple-choice tests of listening comprehension typically
include tasks such as 1) listening to a sentence and then identifying
the correct paraphrase from several choices, 2) listening to short
dialogues and then finding the correct choice to a question about the
dialogue, and 3) listening to a short talk and then answering
comprehension questions based on that talk. The propositional content
of this type of test is typically restricted to academic topics, while
the illocutionary acts are typically ideational and manipulative.
In the paraphrase item type two basic tasks are required: 1)
comprehending a single spoken sentence (stem) and 2) recognizing the
correct paraphrase of this sentence (key) from among four written
sentences. The majority of the items of this type require only
grammatical competence for successful completion. Most depend
primarily on the knowledge of lexical signification, or of the
propositional content expressed by syntactic structure. Furthermore,
these items can be regarded as context-reduced, in that they are
generally unconnected with each other and their references are to
fictitious persons, objects, place and actions. In general, the task
of recognizing paraphrases is an extremely artificial one and requires
virtually no communicative performance, in that this task focuses
exclusively on propositional signification. In addition, the lack of
context renders this item type highly artificial.

In the lecture, or short talk test type there is generally a
variety of discourse organization, including generalization and
development, as well as a variety of illocutionary acts and
linguistic forms. The basic tasks in this test type are 1)
comprehending a spoken discourse and 2) answering direct information
questions based on that discourse. In this type of test, the extent
of the discourse is much more substantial than that in the paraphrase
type. The context is also much more extensive. Unfortunately,
however, the lectures are frequently highly artificial, in that they
sound like "read" presentations and fail to include the kinds of
hesitations and restatements that characterize oral discourse. There
is little challenge to the test taker to interact with the text and
consequently little opportunity for authentic language use.

3.1.2 Multiple-choice tests of structure

The language performance on this type of test is typically
restricted entirely to single sentences, and thus has little potential
for involving discourse. There is generally a variety of propositional
content, while the illocutionary force is typically restricted to the

ideational function and the form to declarative sentences. In this test type the basic task is to recognize the syntactic form that will correctly complete an incomplete statement. The items in this type of test are generally context-reduced, in that they represent isolated propositions, although there is generally some attempt to contextualize them. Frequently, however, this context is totally irrelevant to the task posed by the item. Consider the following item, for example:

> The first ornithischia appeared on the Earth
> _____ the early Mesozoic era, some 200
> million years ago.
>
> (A) when
> (B) or
> (C) and
> (D) during

This statement, if found in an authentic discourse, would presuppose that the reader is familiar with the terms "ornithischia" and "Mesozoic era". In the test context, however, this information is irrelevant to the syntactic structure that requires the preposition "during". If the test taker attempts to process this sentence as an authentic use of language and is not familiar with these terms, the item is context-reduced and may be more difficult than if the meanings of these terms were ignored entirely. Because of the unnecessarily difficult context, items such as these probably engage other competencies even though they are intended only to measure grammatical competence.

3.1.3 Multiple-choice tests of reading comprehension

Of the various types of multiple-choice test, the reading comprehension test has, in my opinion, the greatest potential for requiring communicative language performance. This is because it is the least restricted with respect to organization of discourse, propositional content, illocutionary force, and forms. There are basically two tasks in this test type: 1) comprehending a written text and 2) providing requested information based on the content of that text. The questions are generally of two types: incomplete statements and direct information questions. The type of information requested is usually both literal and inferential. Items in this test type may measure grammatical, cohesive, and illocutionary competence. Strategic competence, to the extent that this is involved in inference and drawing on relevant extra-textual knowledge, can also be measured, should this be desirable.

In general, while multiple-choice tests are highly restricted in terms of the type of performance required, I believe they can be used effectively to measure the receptive skills of listening and reading, and to measure the full range of competencies required in these two skills.

3.3 Oral interviews

The oral interview is probably the premier "direct" test of language proficiency. It is nearly the opposite of the multiple-choice test, in that it can require authentic language use, or communicative performance. While generally limited to the aural/oral channel, both receptive and productive modes can be measured, as can the full range of competencies involved in the skills of listening and speaking. The extent to which this test achieves its full potential, however, depends on the elicitation and rating procedures. The skillful interviewer will lead the candidate through a variety of topics, elicit a variety of illocutionary acts, and present a variety of contexts. Indeed, in a well-conducted interview, the candidate may nearly forget that it is a test.

All too frequently, however, the candidate's performance is rated solely in terms of grammar, pronunciation, vocabulary, and perhaps fluency. Such a rating scale fails to evaluate aspects of either discourse competence, such as cohesion and rhetorical (conversational) organization, or of sociolinguistic competence, such as appropriateness of register and naturalness.

Another common characteristic of rating scales that have been developed for oral interviews is the definition of the scale points, or levels, in terms of specific contexts and subject matter. A well-known example of this type of scale definition is that of the Interagency Language Roundtable (ILR) oral interview (formerly the Foreign Service Institute (FSI) oral interview). This scale has been adopted and expanded by such diverse organizations as the American Council of Teachers of Foreign Languages (ACTFL), in its "Provisional Proficiency Guidelines" and the Australian Department of Immigration and Ethnic Affairs, in its "Australian Second Language Proficiency Ratings (ASLPR)". This type of scale definition may be quite useful for specific situations such as those of the agencies of the U. S. government, or of modern language departments in U. S. colleges and universities, or of a large-scale migrant program in Australia. Because of the effect of context on communicative language use, however, context dependent definitions limit not only the use of such scales, but more importantly, their interpretation, to the specific situations for which they are designed. Thus, ratings on these different scales are of little use for comparative purposes. How comparable, for example, are the following scale definitions?

> Can handle with confidence but not with facility
> most social situations including introductions and
> casual conversations about current events, as well
> as work, family, and autobiographical information;
> can handle limited work requirements, needing help
> in handling any complications or difficulties.

<div align="right">(Lowe, 1980:1-5)</div>

Can narrate, describe, and explain in past,
present, and future time. Can communicate facts--
what, who, when, where, how much--and can explain
points of view in an uncomplicated fashion, but
cannot conjecture or coherently support an
opinion. Can talk in a general way about topics
of current public interest (e.g., current events,
student rules and regulations), as well as
personal interest (work, leisure time activities)
and can give autobiographical information. Can
make factual comparisons, such as college life vs.
high school life. . . . Can make a point
forcefully and communicate needs and thoughts in a
situation with a complication (e.g. calling a
mechanic for help with a stalled car, explaining
suspicious-looking possessions to a customs
official).'

(ACTFL, 1982:   )

Has restricted register flexibility though, where
a specialist register has been experienced, will
have acquired some features of it. . . . Can give
detailed information about own family, living
conditions, educational background; can describe
and converse on everyday things in his environment
(e.g., his suburb, the weather); . . . can
communicate on the spot with fellow workers or
immediate superior (e.g., ask questions about job,
make complaints about work conditions, time off,
etc.),

(Ingram and Wylie, 1981:14)

The first quotation is from the definition of level two on the ILR
rating scale; the second is from the ACTFL "Provisional Proficiency
Guidelines" definition of the "advanced" level, which is considered
comparable to level two on the ILR scale, and the third is from the
ASLPR definition of level two. While the ILR statement is quite
general, the ACTFL and ASLPR statements are much more specific. The
problem this creates for comparability of ratings is that one is not
certain that interviewers using the different scales would try to
elicit exactly the same language functions (narrating, describing,
explaining, explaining, conjecturing, supporting an opinion, making
factual comparisons, communicating needs and thoughts). Nor is it
likely that the same content areas (work, leisure time activities,
student rules and regulations, college and high school life) or social
registers will be of relevance to American college students,
employees of U. S. government agencies, and immigrants to Australia.
If there is this much potential for difference in elicitation and
interpretation of ratings from scales such as these, which are very
closely related in terms of their development, comparability of
interpretation is even less attainable when quite different types of
scales are examined.

A more generalizeble end interpretable approach to scale definition, I believe, is to define levels on separate scales in terms of the characteristics of the various components of communicative competence. In the Oral Interview Test of Communicative Proficiency in English (Bachman and Palmer, 1983), for example, level two is defined in terms of several components, such as:

Grammar

> Limited range of both morphologic and syntactic structures, but with some evidence of systematic rules; control of some structures used, but with many error types

Cohesion (part of discourse competence)

> Moderate cohesion, including coordination and simple subordination; sometimes confusing relationships among ideas

Sensitivity to Register (part of sociolinguistic competence)

> Evidence of two registers and control of either formal or informal register

<div align="right">(Bachman and Palmer, 1983:2-4)</div>

Interviewers are instructed to elicit topics, illocutionary acts, and sociolinguistic registers appropriate to the context and to the candidate's needs and interests. Thus these factors do have an important effect on the communicative performance elicited in the interview. But since the scale definitions are independent of context and subject matter, the interviewer is not constrained to elicit a particular set of discrete grammatical structures, vocabulary items, or speech acts. This is not to claim, however, that defining such scales is not problematical. On the contrary, the identification and ranking of illocutionary acts in terms of appropriateness and level, for example, is extremely complex. Nevertheless, I believe that this approach to scale definition has a great deal of potential for providing a "common yardstick" for rating any given speech sample in terms of the components of communicative language proficiency.

3.3 Cloze tests

The cloze continues to be an enigma. While it appears to approximate quite closely the kind of processing involved in reading, and thus to involve communicative performance, it nevertheless is generally perceived by test takers as a highly artificial task. Indeed, much of the research with variations in this procedure has been motivated, in part at least, by the desire to overcome its lack of appearance of validity. I believe that this perceived artificiality is largely a function of the random deletion procedure, which frequently results in items that are nearly impossible to complete. Perhaps time and conditioning will be the ultimate solution to this problem. After all, even the multiple-choice test, which is

BEST COPY AVAILABLE

now widely accepted by test takers as the prototypical test, was once
perceived as totally lacking in "face validity".

From my own research with the cloze I am convinced that it can
measure the full range of competencies involved in reading. The key to
this potential, however, lies in the specific words that are deleted,
and to assure that the specific competencies one wishes to measure are
in fact measured, it is essential to abandon the random deletion
procedure for a rational one in which the test developer selects the
words to be deleted according to criteria defined in the content
specifications of the test.

4   Implications for measurement theory

Given the range of language performance required and the
competencies measured by language proficiency tests, it would seem
useful to consider the extent to which such tests can be adequately
analyzed by current psychometric theory. One assumption of test
theory, both classical true-score and latent-trait models, is that
test items are locally independent (Rasch, 1980; Lord and Novick,
1968). This means that the probability of an individual's answering an
item correct is a function only of his or her ability level and the
difficulty level of that single item. For this assumption to be
satisfied, test developers must write and arrange items so that they
are as independent of each other as possible in terms of the tasks
required and content included.   This is clearly at odds with
communicative language performance, in which the "items" of discourse
are by definition related to each other and to a given context. (It
should be noted that Holland (1981) has suggested a less restrictive
assumption, that of "local nonnegative dependence", that may provide a
practical means for determining whether item response models fit a
given set of item data.)

A second assumption of currently available latent-trait models is
that the test items comprise a unidimensional scale, that is, that
they all measure a single trait or ability (Lord and Novick, 1968;
Lord, 1980). This assumption would also appear untenable, not only in
terms of current theories of language proficiency, but also in light
of recent research in language testing, which indicates that language
proficiency is multidimensional (e.g., Swinton and Powers, 1980;
Bachman and Palmer, 1981, 1982; Dunbar, 1982; Carroll, 1983; Upshur
and Homburg, 1983). As with the assumption of local independence,
attempts by test developers to satisfy the assumption of
unidimensionality as well result in items that are artificially
restricted in their form and content. In fact, the quintessential
"discrete-point" item might be regarded as unidimensional.

If current theoretical frameworks and research describe
communicative language proficiency as comprising several distinct but
related traits, and communicative language performance as occurring in
the context of discourse, with interrelated illocutionary acts
expressed in a variety of forms, it would seem that language tests
would provide both a challenge and an opportunity for psychometricians
to test the assumptions of current models and to develop more powerful
models if necessary.

5   Implications for test development

There would appear to exist a similar challenge and opportunity
for test developers to find more creative test procedures and formats.
One such procedure, a variation of the dictation called the "copy-
test" (Cziko & Lin, 1984), involves the visual presentation of
material, and has considerable potertial as measure of text
processing.   It is not unreasonable to expect that advances in
microcomputer technology, along with its increasing availability, will
provide the means for making this testing technique feasible for
large-scale testing in the next few years.

Within the multiple-choice framework, I believe it would be
useful to experiment further with items in which some of the
distractors are partially correct. For example, the key response
would be completely correct in terms of syntax, cohesion, coherence,
and perhaps register, while the distractors might be syntactically
current, but not cohesive, syntactically and cohesively correct, but
not in the appropriate register, and so forth. From items such as
these it might be possible to derive scores for these different
aspects of communicative competence. This type of item has been
examined by Fashady (1980).

6   Conclusion

In this paper I have presented a framework for examining
performance on language tests and have attempted to demonstrate how
this framework might provide some insight into the types of language
performance elicited and the language competencies measured by such
tests.   At this point I would like to venture some opinions regarding
the extent to which language tests can or must comprise measures of
communicative competence. First, I believe that it is possible for
tests that do not involve communicative performance to measure some
aspects of communicative competence. Second, it is quite possible
that not all the traits of communicative competence are equally
relevant to the language use needs of a given group. The ability to
use language to perform imaginative functions, for example, is
probably of less importance to college students, unfortunately, than
the ability to perform ideational functions such as defining,
describing or arguing. Finally, it may well be that not all the
relevant abilities of communicative competence are measurable within
the limitations of any given testing program.

A genuine commitment on the part of language testing researchers
and language test developers to content and construct validity
requires the constant re-examination of the objectives of our tests
and a reassessment of the techniques we employ for eliciting
communicative language performance. The innovations that result from
this re-examination will have implications for both test development
and test theory, in that they will require creative applications of
current models and technology, and may stimulate the creation of new
models and new technology.

I disagree with those who feel that the most pressing need is for further development of theoretical models. I believe that current models are sufficiently well-defined to permit empirical verification. I also disagree with those who disparage ever being able to characterize all the components of communicative competence, for such a position would lead us to cease empirical research. The highest priority, I believe, is for us to attempt to move to the paradigmatic stage of scientific development and begin empirical verification of a current theoretical model of communicative language proficiency. Furthermore, I believe that the most effective means of such empirical verification is through language testing research.

Bibliography

American Council of Teachers of Foreign Languages (ACTFL). 1982. ACTFL provisional proficiency guidelines. Hastings-on-Hudson, New York: ACTFL.

Bachman, L. F. and A. S. Palmer. 1981. "The Construct Validation of the FSI Oral Interview". Language Learning 31, 1:67-86.

   - 1982. "The Construct Validation of some Components of Communicative Proficiency". TESOL Quarterly 16, 4:449-65.

   - 1983. Oral interview test of communicative proficiency in English. (Manuscript).

   - 1984. "Some Comments on the Terminology of Language Testing". In Rivera, C. (ed). Communicative competence approaches to language proficiency assessment: Research and applications. Clevedon, England: Multilingual Matters.

   - Forthcoming. Basic concerns in language test validation. Reading, Mass.: Addison-Wesley.

Bialystok, E. and E. Ryan. Forthcoming. "A Metacognitive Framework for the Development of First and Second Language Skills". In Forrest-Pressley, D. L., G. E. MacKinnon and T. G. Waller, (eds). Meta-cognition, cognition and human performance. New York: Academic Press.

Campbell, D. T. and D. W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix". Psychological Bulletin 56, 2:81-105.

Canale, M. and M. Swain. 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing". Applied Linguistics 1, 1:1-47.

Carroll, B. J. 1980. Testing communicative performance: an study. Oxford: Pergamon Press.

2

Carroll, J. B., 1961. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students". In Center for Applied Linguistics. Testing the English proficiency of foreign students. Washington, D. C.: Center for Applied Linguistics.

- 1983. "Psychometric Theory and Language Testing". In Oller, J. W. (ed). Issues in language testing research. Rowley, Mass.: Newbury House.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge, Mass.: Massachusetts Institute of Technology Press.

Cziko, G. and J. Lin. 1984. "The Construction and Analysis of Short Scales of Language Proficiency: Classical Psychometric, Latent-trait, and Nonparametric Approaches". TESOL Quarterly 18, 4:627-47.

Dunbar, S. B. 1982. "Construct Validity and the Internal Structure of a Foreign Language Test for Several Native Language Groups". Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Faerch, C. and G. Kasper. 1983. "Plans and Strategies in Foreign Language Communication". In Faerch. C. and G. Kasper (eds.) Strategies in interlanguage communication. London: Longman.

Farhady, H. 1980. Justification, development, and validation of functional language tests. Unpublished Ph.D. dissertation, University of California at Los Angeles.

Halliday, M. A. K. 1976. "The Form of a Functional Grammar". In Kress, G. (ed). Halliday: System and function in language. Oxford: Oxford University Press.

Holland, P. W. 1981. "When Are Item Response Models Consistent with Observed Data?" Psychometrika 46, 1:79-92.

Hymes, D. H. 1972. Towards communicative competence. Philadelphia: University of Pennsylvania Press.

Ingram, D. E. and E. Wylis. 1981. Australian second language proficiency ratings (ASLPR). Brisbane: Australian Department of Immigration and Ethnic Affairs.

Lado, R. 1961. Language testing: The construction and use of foreign language tests. London: Longman, Green & Company.

Lord, F. M. 1980. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. and M. R. Novick. 1968. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.

Lowe, P. 1980. manual for LS oral interview workshops. Washington, O.C.: Language School, CIA.

Munby, J. 1978. Communicative syllabus design: A sociolinguistic model for defining the content of purpose-specific language programmes. Cambridge: Cambridge University Press.

Rasch, G. 1960. Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute for Educational Research.

Savignon, S. J. 1983. Communicative Competence: Theory and classroom practice. Reading, Mass.: Addison-Wesley.

Swinton, S. S. and D. E. Powers. 1980. Factor analysis of the test of English as a foreign language for several language groups. TOEFL Research Reports, Report 6. Princeton, N. J.: Educational Testing Service.

Upshur, J. A. and T. J. Homburg. 1983. "Some Relations among Language Tests at Successive Ability Levels". In Oller, J. W. (ed). Issues in language testing research. Rowley, Mass.: Newbury House.

Widdowson, H. G. 1978. Teaching language as communication. Oxford: Oxford University Press.