

*The Construct Validation of Some Components of Communicative Proficiency**

Lyle F. Bachman and Adrian S. Palmer

The notion of communicative competence has received wide attention in the past few years, and numerous attempts have been made to define it. Canale and Swain (1980) have reviewed these attempts and have developed a framework which defines several hypothesized components of communicative competence and makes the implicit claim that tests of components of communicative competence measure different abilities. In this study we examine the construct validity of some tests of components communicative competence and of a hypothesized model. Three distinct traits—linguistic competence, pragmatic competence and sociolinguistic competence—were posited as components of communicative competence.

A multitrait-multimethod design was used, in which each of the three hypothesized traits was tested using four methods: an oral interview, a writing sample, a multiple-choice test and a self-rating. The subjects were 116 adult non-native speakers of English from various language and language-learning backgrounds. Confirmatory factor analysis was used to examine the plausibility of several causal models, involving from one to three trait factors. The results indicate that the model which best fits the data includes a general and two specific trait factors—grammatical/pragmatic competence and sociolinguistic competence. The relative importance of the trait and method factors in the various tests used is also indicated.

Our understanding of language proficiency has been considerably broadened in the past few years by the notion of communicative competence, and frameworks of language use which have been proposed recently (Canale & Swain 1980, Munby 1978) differ considerably from those advocated earlier by Lado (1961), Carroll (1968) and other language testing specialists. Canale and Swain (1980) reviewed numerous attempts to define communicative competence and developed a framework which not only defines several hypothesized components of communicative competence but also makes the implicit claim that components of communicative competence comprise dis-

Mr. Bachman, Assistant Professor of English as a Second Language at the University of Illinois, teaches courses in language testing and research design.

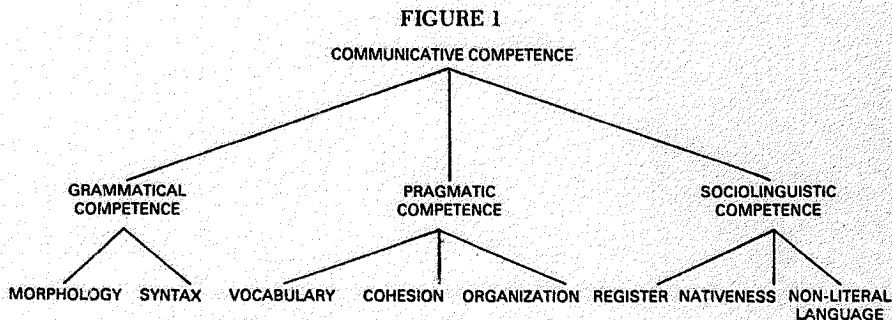
Mr. Palmer, Associate Professor of English at the University of Utah, teaches courses in ESL methodology, theory, and language testing.

*Primary funding for this research was provided by the University of Utah Research Fund. The Division of English as a Second Language, University of Illinois provided released time and research assistance.

tinct underlying abilities. In their framework the specific abilities are: 1) grammatical competence, which includes lexicon, morphology, syntax, sentence-level meaning and phonology; 2) sociolinguistic competence, which includes sociocultural appropriateness rules and discourse rules; and 3) strategic competence, which comprises various verbal and nonverbal communication strategies which are employed to compensate for deficiencies in (grammatical and sociolinguistic) competence or to accommodate the vicissitudes of the communication situation.

Recent efforts to develop tests based on "communicative" frameworks have been extensive, ranging from tests for general use in minority language settings (Canale 1981), to tests or test batteries designed for specific language use situations (Carroll 1980, Farhady 1981), to tests for specific types of language programs (Morrow 1977). But while theorizing and test construction continue apace, little empirical research is available regarding the validity of these theoretical frameworks or their components as psychological constructs. In this study, we have begun an empirical investigation of communicative proficiency which both draws from and, we believe, clarifies current theoretical frameworks.

The theoretical framework which this study examines comprises three main components, or traits: grammatical competence, pragmatic competence and sociolinguistic competence. *Grammatical competence* includes morphology and syntax, both of which vary in range and accuracy. Phonology and graphology are not included here because we view these more as channels than as components. This is because there appears to be a critical level of pronunciation accuracy (or legibility), below which verbal communication completely breaks down, while above that level, communicative language use is possible, with increasing facility as accuracy increases. Below this "threshold" level required for communication, therefore, it is not possible to implement the components of competence. (See Bachman & Palmer 1981c for further discussion.) *Pragmatic competence*, which we associate with the ability to express and comprehend messages, includes the sub-traits vocabulary, cohesion and organization or coherence. Our inclusion of vocabulary as a sub-trait of pragmatic, rather than grammatical, competence is based on our frequent observation that certain non-native speakers with little or no grammatical competence are nevertheless able to maintain some meaningful communication on the basis of their knowledge of vocabulary alone. At the other end of the continuum, among speakers whose grammatical competence is virtually complete, those who possess extensive vocabularies are able to express a greater variety of messages with greater precision and efficiency than are those with only moderate vocabularies. *Sociolinguistic competence* includes the following sub-traits: distinguishing of registers, nativeness, and control of non-literal, figurative language and relevant cultural allusions. The framework of main and sub-traits used in this study is presented in Figure 1 below.



While this model makes specific claims regarding the different abilities or constructs which comprise communicative competence, there is a considerable body of research which suggests the presence of a "general" construct of language competence. Oller (1976, 1979) and others originally found that a general factor apparently accounted for the greatest proportion of reliable variance in language test scores and concluded that language proficiency is essentially a single unitary trait, or ability. More recently, however, several researchers (Carroll 1980, Bachman & Palmer 1981a, Upshur & Homburg 1980) have demonstrated that the unitary trait hypothesis is not supported empirically, and suggest that models which include both a general trait and one or more specific traits provide better explanations for language test data.¹ As Oller (1981, forthcoming) has indicated, there now seems to be a consensus among researchers that models including both general and specific factors will provide the best explanations for language test data. Therefore, in addition to examining the nature of specific components of communicative competence, it was also of interest in this study to determine the extent to which a general trait is part of communicative competence.

Construct Validation

Language proficiency tests have typically been developed to reflect specific abilities or constructs, often with the choice of testing method of secondary priority and made on the basis of practical considerations. Several recent studies, however, (Clifford 1978, 1980, Corrigan & Upshur 1978, Brütsch 1979, Bachman & Palmer 1981a), have demonstrated the influence of test method on test scores. These studies, as well as the present one, employ the multitrait-multimethod (MTMM) matrix as a construct validation paradigm (Campbell & Fiske 1959). Recognizing that the variance on a test is comprised not simply of "true" variance and random error variance, as in classical test theory, MTMM studies distinguish *two* sources of error variance: that due to test method and that due to random measurement error. In a test which measures only one construct, the total variance (ST^2) can be

¹ As Carroll (1980) has pointed out, however, this constitutes a reaffirmation, rather than the discovery of two-factor models as explanations for language test scores.

analyzed into three components: "true" variance (S_t^2), or that due to the construct or trait tested, method variance (S_m^2) and error variance (S_e^2):

$$S_T^2 = S_t^2 + S_m^2 + S_e^2$$

An extension of this analysis applies to tests which measure more than one trait, where the "true" variance may be further broken down into that due to the different traits tested ($S_{t_1}^2 \dots S_{t_n}^2$):²

$$S_T^2 = S_{t_1}^2 + S_{t_2}^2 \dots S_{t_n}^2 + S_m^2 + S_e^2$$

Since the purpose of construct validation is to determine "what constructs account for variance in test performance" (Cronbach and Meehl, 1955), the logic of construct validation requires that the variances due to the construct(s) under consideration be distinguishable, and that these sources of variance also be distinguishable from error variance (both method and measurement error.)

In order to make possible these distinctions, MTMM studies employ several methods for testing each of several putatively distinct traits. In the present study, competence in each of the three hypothesized traits of communicative proficiency was measured by four methods: an interview, a writing sample, a multiple-choice test and a self-rating. The combination of these three traits and four methods provides for twelve trait-method units, or tests, each a combination of a single trait with a single method. These twelve measures (X_1 to X_{12}) are represented in Figure 2.

FIGURE 2
Multitrait-Multimethod Design
For Study

METHODS MAIN TRAITS	ORAL INTERVIEW	WRITING SAMPLE	WRITTEN OBJECTIVE MULTIPLE-CHOICE	SELF-RATING
GRAMMATICAL COMPETENCE	(X ₁) Portion Rated For Grammatical Competence	(X ₂) Portion Evaluated For Grammatical Competence	(X ₃) Test of Grammatical Competence	(X ₄) 3 Range Questions 3 Accuracy Questions
PHRAGMATIC COMPETENCE	(X ₅) Portion Rated For Pragmatic Competence	(X ₆) Portion Evaluated For Pragmatic Competence	(X ₇) Tests of Pragmatic Competence	(X ₈) 3 Vocabulary Questions 3 Cohesion Questions 3 Organization Questions
SOCIO- LINGUISTIC COMPETENCE	(X ₉) Portion Rated For Socio- linguistic Competence	(X ₁₀) Portion Evaluated For Sociolinguistic Competence	(X ₁₁) Tests of Sociolinguistic Competence	(X ₁₂) 3 Register Questions 3 Nativeness Questions 3 Non-literal language Questions

² While much of psychometric theory and methodology is aimed at tests which are monotonic, or which measure a single trait, so-called "integrative" language tests such as oral interviews, which, in light of the conclusions reached in this study, would appear to measure more than one trait, suggest this extension of the single trait model.

Instrumentation

Interview method. The oral interview used was an adaptation of the FSI oral interview. Elicitation procedures and rating scales grew out of a prior study and were further developed specifically to elicit a ratable sample of each of the hypothesized traits. The interview was conducted by two examiners, and required approximately 25 minutes to complete. In conducting the interview, one interviewer dressed formally, and assumed a formal role, while the other assumed an informal role. The room was arranged with the two interviewers in separate areas, with the candidate seated so that attention would be directed towards only one interviewer at a time.

The interview began with a short greeting and introduction by the informal interviewer, followed by a short "warm-up" conversation which was also used to make an initial estimation of the candidate's ability. During this conversation, appropriate topics for further discussion were identified. Except in cases of very low proficiency, the candidate was asked to prepare a short formal presentation on this topic, and given a few minutes to prepare.

The candidate was then introduced to the formal interviewer, who was given a fictitious role appropriate to the candidate's interests and ability (for example, immigration official, college admission counsellor, or a prospective employer). The candidate was then interviewed by the formal interviewer. This interview included the short oral presentation which he/she had prepared.

At the end of this part of the interview, the candidate returned to the informal interviewer, who asked him/her to participate in a role-play in which he/she was to imagine that the informal interviewer was his/her best friend. The role-play required calling the "friend" on the phone, inviting him/her to go out somewhere, and making arrangements for when and where to meet, and for transportation. This role-play was followed by a brief informal "wind-up" and leave-taking. (See Bachman & Palmer, 1981b for further discussion of this test.)

Writing sample method. The writing sample test consisted of several subsections requiring a variety of writing tasks, ranging from short answers to more extensive composition. In one section, for example, subjects were given a picture, instructed to describe it in as much detail as possible and told that they would be graded on vocabulary. In another section they were asked to provide appropriate greetings and closings for a business letter, a note to a professional colleague, a note to a newspaper carrier, and a love letter. The text for each of these letters or notes was provided on authentic stationary or note paper. In another section they were asked to write a short composition comparing and contrasting the place in which they grew up with the place where they were currently living.

Multiple-choice method. The multiple-choice method consisted of 119 5-choice items in 10 sub-tests. Item stimuli included pictures, short sentences, dialogues and short paragraphs.

The following items exemplify this test:

"So your motorcycle race was fun, huh?"

"Yeah, it was (1) *really something*. I'll tell you all about it when (2) *we meet* later."

"O.K. (3) *See you* at the pool."

"Yeah. Be ready to (4) *get beat!*"

"(5) *Fat chance!*"

(sensitivity to register)

I went into town and _____

(1) got lost.

(2) got off the way.

(3) got off my way.

(4) lost the path.

(5) lost my path.

(nativeness)

"He's pretty absent-minded, don't you think?"

"He sure is. And he's not even a _____."

(1) hockey player

(2) college professor

(3) sales person

(4) big executive

(5) secretary

(cultural stereotypes)

Self-rating method. (See Bachman & Palmer 1982 for further discussion of this test.) The self-rating method consisted of 24 items using 4-point Likert-type scales. These items required the respondents to make three types of self-ratings:

1) ratings of their ability in the trait areas, for example:

"How many English words do you know?"

Very few

Several
hundred

A lot, but not
as many as most
Americans know

As many as most
Americans know

(Pragmatic-vocabulary)

2) ratings of their difficulty or deficiency in the trait areas, for example:

"How hard is it for you to put several English sentences together in a row?"

Impossible

Very hard

Not very hard

Very easy

(Pragmatic-cohesion)

3) ratings of their ability to recognize the presence or absence of the traits, for example:

"In general, can you tell when someone makes a grammatical mistake?"

No, almost
never

Sometimes

Usually

Yes, almost
always

(Grammatical-accuracy)

All of these tests were pre-tested with non-native English speaking students at the University of Illinois. Interview procedures and scoring criteria were simplified and clarified, and the other tests were revised—shortened and made less ambiguous—on the basis of the pre-test results.

In addition to these tests, subjects completed a questionnaire which included questions regarding the conditions under which they learned and use English, their reasons for learning English and various types of demographic information.

Subjects

The subjects were 116 non-native English speakers from the Salt Lake City area ranging in age from 17-67, with a median age of 23. They were from 36 different countries, with 18 native language backgrounds, and had lived in the U.S. from a few months to over 10 years, with a median length of stay in the U.S. of 1.2 years. 75 had studied English for at least 1 year in their native country, while 41 had not studied English in their native country at all. In addition, only half had studied English for 1 year or more in the U.S. 72 were university students (64 undergraduate, 8 graduate) majoring in 33 different fields, 20 were in adult education programs, 10 were studying English intensively, 2 were high school students, while 12 were not students (4 of these were university faculty members.) 49 indicated that they knew a foreign language other than English. Of these, 11 indicated a better knowledge of this language than of English.

Procedures

Administration. Each subject took all tests within a two-week period. All subjects completed the self-rating first in order to prevent their responses from being influenced by their performance on the other tests. The order in which the subjects completed the remaining three tests varied due to the need to administer the writing sample and multiple choice tests to most of the subjects in groups during regularly scheduled class periods while scheduling these same subjects for individual interview tests over the entire two-week period. Tests were administered by the following personnel. Group tests were administered by the subjects' classroom teachers (except for a few of the highest level subjects—University of Utah staff members—who completed their tests on their own.) The majority of the interview tests (80) were administered by the authors (Bachman and Palmer); The remainder were administered by Palmer and George Trosper, a research assistant working on the project. The final two tests were administered by Bachman and Fred Davidson at the University of Illinois.

Scoring. At the conclusion of each oral interview test, the two interviewers individually assigned the subject ratings on each of the sub-traits using the sub-trait definitions given in Figures 3-5.

FIGURE 3
Main trait: Grammatical Competence

Main trait Rating	Range	Subtraits	Accuracy
0	No systematic evidence of morphologic and syntactic structures.		Control of few or no structures. Errors of all or most possible types frequent.
1	Limited range of both morphologic and syntactic structures, but with some systematic evidence.		
2		Control of some structures used,	
3	Large, but not complete, range of both morphologic and syntactic structures.		but with many error types.
4			
5	Complete range of morphologic and syntactic structures.		Control of most structures used with few error types.
6			No errors not expectable of a native speaker.

FIGURE 4
Main trait: Pragmatic Competence

Main trait Rating	Vocabulary	Subtrait's Cohesion	Organization
0	Limited Vocabulary (A few words and formulaic phrases)	Language completely disjointed	Natural organization only (i.e., not consciously imposed)
1	Small Vocabulary	Very little cohesion; relationships between structures not adequately marked.	or Poor ability to organize consciously
2	Vocabulary of moderate size	Moderate cohesion, including coordination.	
3	Large Vocabulary	Good cohesion, including subordination.	Moderate ability to organize consciously
4	Extensive Vocabulary	Excellent cohesion, using a variety of appropriate devices.	Excellent ability to organize consciously

FIGURE 5
Main trait: Sociolinguistic Competence

Main trait Rating	Distinction of Formulaic	Subtraits		Cultural References
		Register Substantive	Nativeness	
0	Evidence of only one register	Evidence of only one register	Frequent non-native but grammatical structures	No Control
1	Evidence of two registers		or Impossible to judge because of interference from other factors	Some Control
2	Evidence of two registers and control of formal or informal	Evidence of two registers	Rare non-native but grammatical structures	
3	Control of both formal and informal registers	Evidence of two registers and control of formal or informal		No non-native but grammatical structures
4		Control of both formal and informal registers		

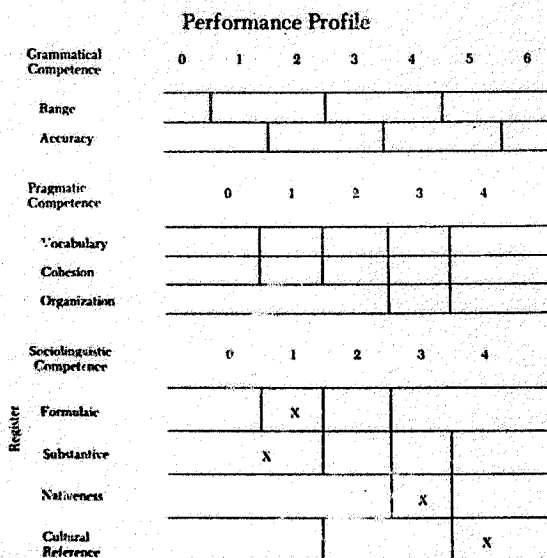
Each interviewer then combined his sub-trait ratings using a non-compensatory system. Subjects were given main trait ratings equal to the highest level at which the criteria were met on all sub-traits. For example, if a subject received a subtrait rating of '3' on vocabulary and organization, but a rating of only '2' on cohesion, his/her main trait rating for pragmatic competence would be '2'. A sample rating protocol (performance profile) is provided in Figure 6, with a hypothetical subject's ratings on sociolinguistic competence filled in. X's indicate the subject's ratings on the subtraits of sociolinguistic competence, while the circled number indicates his non-compensatorily derived main trait rating. The main trait ratings of the two interviewers were averaged for the statistical analysis of the data.

The writing sample tests were scored by the researchers (Bachman and Palmer) in essentially the same manner as the oral interview tests—using the same definitions and performance profile sheets.

The multiple-choice and self-rating tests were objectively scored. The range of possible scores on each sub-trait's portion of these tests was divided into intervals corresponding to the intervals on the performance profile sheets, and sub-trait and main trait scores were assigned accordingly. (See Bachman and Palmer (forthcoming) for the details.)

Analyses. Distributions, correlations and reliabilities were computed using SPSS Version 8 (Nie *et al.* (1975, Hull & Nie 1979), and maximum likelihood confirmatory factor analyses were computed using LISREL 4 (Jöreskog & Sörbom 1978), both on the CYBER system at the University of Illinois.

FIGURE 6



Results

Reliabilities. Since reliability is a requisite for validity, reliability estimates were computed for all the measures in the study. In addition to alpha coefficients, which were computed for all measures, inter-rater correlations were computed for the interview and writing methods. The obtained reliability estimates are given in Table 1. These range from .747 to .980, indicating that all the measures are of acceptable reliability.

Confirmatory Factor Analysis. Confirmatory factor analysis (CFA) is a technique for statistically evaluating or testing the extent to which a given set of hypotheses are confirmed by the relationships observed in a body of data. In stating these hypotheses explicitly, the researcher posits a model which specifies the number of factors which underly the measured variables, as well as the exact relationships between the factors and the measured variables, and among the factors themselves. This causal model then constitutes a theoretical explanation for the correlations actually obtained in the data. The extent to which this explanation provides a statistically significant fit to the data can be determined by the chi-square (χ^2) statistic. Likewise, the relative fit of different causal models can be compared by examining differences among their chi-squares. In general, the *smaller* the chi-square, relative to its associated degrees of freedom, the better the fit.³

³ Chi square values and probability levels for testing hypotheses with confirmatory factor analysis are interpreted exactly opposite to the way they are used in conventional hypothesis testing. This is because we are expressing the probability that the hypothesized model can be accepted as an explanation of the data, and not the probability that some null hypothesis can be rejected, as in classical hypothesis testing.

TABLE 1
Reliabilities of Measures

	Inter-rater	Alpha
Interview:		
Grammar	.925	.980
Pragmatic	.890	.965
Sociolinguistic	.906	.951
Writing Sample:		
Grammar	.912	.966
Pragmatic	.923	.934
Sociolinguistic	.861	.942
Multiple-choice:		
Grammar		.972
Pragmatic		.747
Sociolinguistic		.839
Self-Rating:		
Grammar		.857
Pragmatic		.905
Sociolinguistic		.917

In multitrait-multimethod studies employing CFA, two types of factors, trait and method, are posited. In this way, it is possible to distinguish trait from method factors, and to clearly distinguish method variance from measurement error. The use of CFA in such studies also permits, indeed requires, the researcher to make absolutely explicit any assumptions or hypotheses regarding correlations among traits or methods, or between traits and methods. The multitrait-multimethod correlations obtained in this study are given in Appendix A.

Models Examined. Although both the unitary and completely divisible trait hypotheses have been rejected by earlier studies, it was nevertheless of interest to test these hypotheses in the present study because the constructs investigated differ considerably from those examined in previous research. The chi-squares, associated degrees of freedom and probability levels for the models representing these two hypotheses are given in Table 2.

TABLE 2

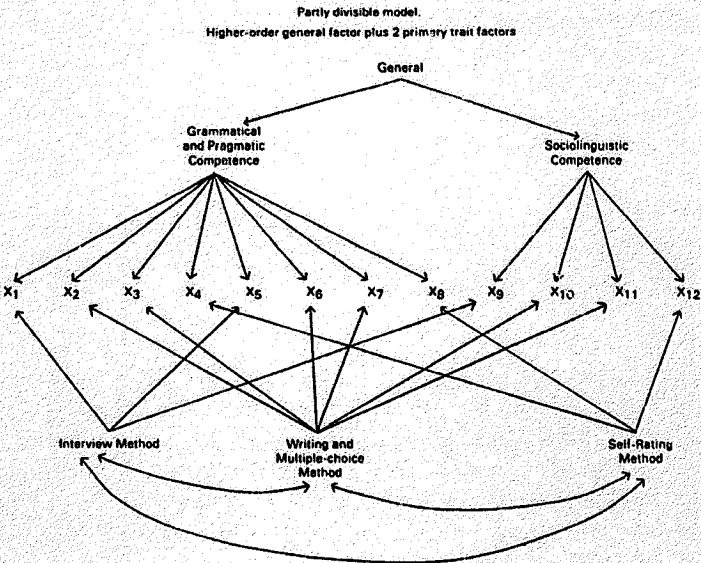
	χ^2	df	p
Completely unitary	86.85†	42	.001
Completely divisible	340.51	42	.0000

†model not identified

As can be seen from the size of the chi-squares and the probability levels, neither of these models can be accepted as an explanation of the data.

Subsequently, several "partly divisible" models of communicative competence were tested. It was initially assumed that there would be no interaction between traits and methods and that the four methods would be uncorrelated with each other. In testing these assumptions, however, we found that partly divisible models with *three* correlated methods in general fit better than models with four uncorrelated methods, and so all of the models discussed below include three correlated method factors. Of the partly divisible models tested, it was found that a model with three correlated trait factors (grammatical, pragmatic and sociolinguistic competence) provided a barely significant fit ($\chi^2 = 48.35$, $df = 36$, $p = .082$), while a model with a general factor and three uncorrelated trait factors provided a better fit ($\chi^2 = 32.83$, $df = 27$, $p = .203$) but which was technically unacceptable.⁴ Of the models tested, that which provided the best fit ($\chi^2 = 30.63$, $df = 27$, $p = .286$) to the data included a higher-order general factor and two uncorrelated primary trait factors. The path diagram which illustrates the relationships among the factors and the observed variables in this model is given in Figure 7. The relative importance of each of the lines or paths in this diagram can be estimated, in the form of factor loadings and factor correlations. The factor loadings and factor correlations are given in Table 3.

Figure 7



⁴ A solution was not reached by the iterative procedure, several of the parameter estimates (factor loadings and uniquenesses) were uninterpretable, and many of the standard errors of estimate were larger than the parameter estimates themselves.

TABLE 3⁵

Factor loadings for Partly Divisible Model: General Factor plus Two Uncorrelated Trait Factors
($\chi^2 = 30.63$, $df = 27$, $p = .286$)

Measure	General	Gram/Prag	Socio	Interview	Writ/M-C	Self-R	Specificity
GRAM INT (X1)	.889	.145	.000	.265	.000	.000	.101
GRAM WRIT (X2)	.859	.285	.000	.000	.236	.000	.092
GRAM M-C (X3)	.621	.766	.000	.000	.350	.000	.076
GRAM SELF (X4)	.726	.051	.000	.000	.000	.328	.290
PRAG INT (X5)	.816	.185	.000	.380	.000	.000	.125
PRAG WRIT (X6)	.856	.164	.000	.000	.088	.000	.166
PRAG M-C (X7)	.538	.396	.000	.000	.004	.000	.301
PRAG SELF (X8)	.757	.090	.000	.000	.000	.597	.031
SOCIO INT (X9)	.742	.000	.219	.311	.000	.000	.250
SICIO WRIT (X10)	.865	.000	.326	.000	.056	.000	.085
SOCIO M-C (X11)	.678	.000	.267	.000	.002	.000	.308
SOCIO SELF (X12)	.669	.000	.167	.000	.000	.596	.080

Method Factor Correlations

	Interview	Writing/ Multiple-Choice	Self- Rating
Interview	1.000		
Writing/ Multiple-Choice	.013	1.000	
Self- Rating	.473	.465	1.000

⁵ The loadings on the general factor are to be interpreted as functioning *through* the primary trait factors. In addition to the loadings on the general, trait and method factors, each measure includes specificity. This specificity is that portion of the variance which is due to the particular combination of variables, minus error of measurement. The obtained alpha coefficients were used to determine the measurement error for the variables.

Factor Loadings. All measures except the multiple-choice test of grammar load most heavily on the general factor, with the largest loadings on this factor occurring for measures using the interview and writing methods. Measures using the writing and multiple-choice method consistently load more heavily on trait than method factors. Indeed, the method loadings for measures of pragmatic and sociolinguistic competence using these two methods are smaller than the specificities for these measures. For measures using the interview and self-rating methods, the situation is reversed; measures using these methods consistently load more heavily on method than

trait factors, with two trait loadings (grammar self-rating and pragmatic self-rating) nearly vanishing.

These results indicate the relative effects of the general, trait and method factors in the different types of tests examined. The interview and self-rating tests used in this study appear to consist largely of general factor variance with some noticeable method variance. The writing sample and multiple-choice measures, on the other hand appear to consist primarily of general and trait factor variance, with negligible variance due to method. These findings are not entirely inconsistent with those of an earlier study (Bachman & Palmer 1981a), in which the two measures involving the interview method loaded most heavily on a general factor, but with the interview measure of reading loading more heavily on the method than trait factor. In that study, measures using the self-rating method also consistently loaded most heavily on general and method factors. We would also note that the results of the present study provide indirect support for the general plus specific trait model, rather than the correlated trait model in that study.

Method factors. The correlations among the method factors indicate substantial correlations between the self-rating and the other two methods, while there is virtually no correlation between the interview and writing methods. The low correlation between the interview and writing methods was to be expected, and provides evidence of their distinctness as methods. The correlations between the self-ratings and the other two methods, however, while not high, are somewhat problematical. It may be that by presenting the self-rating questions in such a way as to encourage subjects to consider both aural/oral and visual modes, we created a halo effect with the other methods.

Trait factors. While we had originally posited separate components of grammatical and pragmatic competence, it is perhaps not surprising that these two appear to cluster together and are distinct from sociolinguistic competence. It may well be that grammar and vocabulary underly, or are necessary for cohesion and organization, and that these are all functions of the organizational aspects of language, both structural and semantic. The sociolinguistic subtraits, on the other hand, may be related more to the affective aspect of language. While this study has not found evidence for separate grammatical and discourse subtraits, we nevertheless feel that there is strong theoretical justification for hypothesizing a distinction between the explicit markers of organization—grammar, vocabulary and cohesion—and the organizational strategies of coherence, and intend to examine the relationship between these aspects in subsequent studies.

With regard to the general factor, the fact that the measures with the heaviest loadings on this factor in general employ the interview and writing methods, suggests that this factor may involve information processing in extended discourse.

Conclusion

In this study we have found empirical support for a model which incorporates some of the components which have been posited by current theoretical frameworks of communicative competence. We believe that this empirical validation is significant in that such theoretical frameworks are already influencing current language teaching pedagogy to the extent that they form the basis for several language teaching curricula, both in methods and materials, as well as for language tests.

In addition to distinct trait components, we have also found evidence for a substantial general factor which affects all the measures in the study, while perhaps of more practical interest are the relative effects of the general, trait and method factors evidenced by the different types of tests used in this study.

Methodologically, we feel this study again demonstrates the power of causal modeling with confirmatory factor analysis as a research paradigm for construct validation. We believe that the explicitness with which research hypotheses and assumptions must be stated in such models greatly facilitates both the formulation and the validation of hypotheses regarding the psycholinguistic constructs which underly scores on language tests.

REFERENCES

- Bachman, L.F. and A.S. Palmer. (Forthcoming). *Basic concerns in language test validation*. Reading, Massachusetts: Addison-Wesley.
- Bachman, L.F. and A.S. Palmer, 1981a. The construct validity of the FSI oral interview. *Language Learning* 31, 1:67-86.
- Bachman, L.F. and A.S. Palmer, 1981b. A scoring format for rating components of communicative proficiency in speaking. Paper presented at the Georgetown University Pre-conference on Oral Proficiency Testing, March 19, 1981.
- Bachman, L.F. and A.S. Palmer, 1981c. Some comments on the terminology of language testing. Paper presented at the Language Proficiency Assessment Symposium, Rosslyn, Virginia, March 14-18, 1981.
- Bachman, L.F. and A.S. Palmer, 1982. The construct validity of self-ratings of communicative competence. Paper presented at the 16th Annual TESOL Convention, Honolulu, May 1982.
- Brütsch, S.M. 1979. Convergent/discriminant validation of prospective teacher proficiency in oral and written French by means of the MLA cooperative language proficiency tests, French direct proficiency tests for teachers (TOP and TWP), and self-ratings. Unpublished Ph.D. dissertation, University of Minnesota.
- Canale, Michael. 1981. A communicative approach to language proficiency assessment in a minority setting. Paper presented at the Language Proficiency Assessment Symposium, Rosslyn, Virginia, March 14-18, 1981.
- Canale, M. and M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1:1-47.
- Campbell, D.T. and D.W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56, 2:81-105.
- Carroll, B.J. 1980. *Testing communicative performance*. Oxford: Pergamon Press.
- Carroll, J.B. 1961. The psychology of language testing. In Alan Davies, ed. *Language Testing Symposium*. London: Oxford University Press.

- Carroll, J.B. 1980. Psychometric theory and language testing. Paper presented at the Second International Language Testing Symposium, Darmstadt, West Germany, May 1980.
- Clifford, R.T. 1978. Reliability and validity of language aspects contributing to oral proficiency of prospective teachers of German. In J.L.D. Clark, ed. *Direct testing of speaking proficiency*. Princeton: Educational Testing Service.
- Clifford, R.T. 1980. Convergent and discriminant validation of integrated and unitary language skills: the need for a research model. In A.S. Palmer, P.J.M. Groot and G.A. Tropper, eds. *The validation of oral proficiency tests*. Washington, D.C.: Teachers of English to Speakers of Other Languages.
- Corrigan, A. and J.A. Upshur. 1978. Test method factors in foreign language tests. Paper presented at the 1978 TESOL Convention, Mexico City.
- Cronbach, L.J. and P.E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52, 4:281-302.
- Farhady, H. 1981. Justification, development and validation of functional language testing. Unpublished Ph.D. dissertation, University of California at Los Angeles.
- Hull, C.H. and H.H. Nie. 1979. *SPSS update*. New York: McGraw-Hill.
- Jöreskog, K.G. and O. Sörbom. 1978. *LISREL user's guide to the analysis of linear structural relationships by the method of maximum likelihood, Version IV, Release 2*. Chicago: National Educational Resources, Inc.
- Lado, R. 1961. *Language testing*. New York: McGraw-Hill.
- Morrow, K. 1977. *Techniques of evaluation for a notional syllabus*. London: Royal Society of Arts.
- Munby, J.L. 1978. *Communicative syllabus design*. London: Cambridge University Press.
- Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner and D.H. Bent. 1975. *SPSS: Statistical package for the social sciences*. 2nd Edition. New York: McGraw-Hill.
- Oller, J.W., 1981. Language testing research (1979-80). In R.B. Kaplan, ed. *Annual review of applied linguistics 1980*. Rowley, Mass.: Newbury House.
- Oller, J.W. (forthcoming). A consensus for the 80s? In J.W. Oller, ed. *Issues in language testing research*. Rowley, Mass.: Newbury House.
- Oller, J.W. and F.B. Hinofotis. 1980. Two mutually exclusive hypotheses about second language ability: indivisible or partly divisible competence. In J.W. Oller and K. Perkins, eds. *Research in language testing*. Rowley, Mass.: Newbury House.
- Upshur, J.A. and T.J. Homburg, 1980. Some relations among language tests at successive ability levels. Paper presented at the Second International Language Testing Symposium, Darmstadt, West Germany, May 1980.

APPENDIX A
MTMM Correlations
 (All correlations significant at $p \leq .001$, $df = 109^*$)

	GRAMINT	GRAMWRT	GRAMMLT	GRAMSLF	PRAGINT	PRAGWRT	PRAGMLT	PRAGSLF	SLGINT	SLGWRT	SLGMLT	SLGSLF
GRAMINT	1.000											
GRAMWRT	.801	1.000										
GRAMMLT	.660	.668	1.000									
GRAMSLF	.656	.632	.524	1.000								
PRAGINT	.857	.751	.636	.678	1.000							
PRAGWRT	.754	.667	.435	.617	.659	1.000						
PRAGMLT	.517	.572	.637	.337	.493	.365	1.000					
PRAGSLF	.760	.607	.627	.758	.708	.656	.428	1.000				
SLGINT	.718	.628	.442	.648	.727	.634	.365	.683	1.000			
SLGWRT	.774	.725	.570	.636	.729	.736	.547	.689	.722	1.000		
SLGMLT	.589	.587	.434	.508	.556	.588	.388	.488	.560	.671	1.000	
SLGSLF	.688	.503	.501	.667	.662	.583	.334	.864	.662	.663	.478	1.000

Construct Validation

*List-wise deletion of missing data reduced sample from 116 to 110.