The Notions of Language Proficiency and Language Dimensionality

Suthathip Thirakunkovit
Faculty of Liberal Arts, Mahidol University, Thailand
E-mail: suthathip.thi@mahidol.edu

Abstract

This study is a literature review on the findings of different studies on the dimensionality of language proficiency. It aims to particularly analyze the assumptions and the statistical findings of relevant research since it has always been an endless debate among linguists and language teachers about whether language proficiency is organized as a unitary ability or multiple divisible competences. For this objective, it is necessary to begin this study by defining the notions of language proficiency. By clarifying the theoretical meanings of language proficiency, I investigated how different researchers discuss the relationships between different language sub-skills and the overall language proficiency. Since there was a major shift of interest from discrete-point testing to the testing of integrated skills in 1970s, this has led to debates and discussion about the nature of language proficiency in communication. Even though there are some degrees of consensus on research evidence for different language sub-skills, the results

are still inconclusive or even contradictory. In trying to summarize the findings of many studies, I believe that it is imperative to assess the implications of the findings of these studies in the benefits of language teaching and testing.

Keywords: language proficiency, language dimensionality, communicative competence

Introduction

In second language acquisition research, there has always been a broad interest in the question about how people acquire and learn a second language. However, in the field of second language assessment, the basic questions are usually about how language proficiency can be measured reliably and validly, and how a test can predict the actual performance of the learners. At first glance, these seem to be easy questions, but do we actually understand the meaning of language proficiency? Do we know how different assessment methods are viewed and used to predict the actual language proficiency of the learners?

This study will review the findings of different studies that discuss different models of language proficiency and approaches that are used to measure language abilities. The implication of the arguments discussed in this study is to help language teachers develop appropriate teaching materials and test experts to increase validity and reliability of their tests.

A Historical Review of Models of Language Proficiency

The endless debate about whether language proficiency consists of a unitary factor or a number of unique underlying factors has been around in the field of language assessment for at least five decades. The hypothesis of "unitary competence hypothesis" originally came from a study of John Oller in 1983. Even though some researchers have argued that the interpretations of Oller's study are invalid due to the inappropriateness of his research procedure (Spolsky, 1981), his hypothesis is still very much alive and

has been put forward by many researchers (Carroll, 1961; Hosley & Meredith, 1979; Spearman, 1904; Vollmer & Sang, 1983). Before I get into the discussion of the two competing hypotheses, I will first discuss how different researchers define the notions of language proficiency.

The Notions of Language Proficiency

Several notions of language proficiency have emerged from the research in second language acquisition. Robert Lado (1961) first proposed that language could be conceived as a system that consisted of several distinct components such as phonemes, morphemes, phrases, clauses, and sentences. However, his idea has not been well-supported because it could mislead to the belief that the test construct was made up only of one unique element of the language.

In 1972, Hymes proposed the concept of communicative competence for language teaching and learning. Hymes viewed that a person who learned a second language should have both knowledge about the language and ability to use it appropriately in a given speech community. Based on this notion of language competence, the successful language learning outcomes refer to the ability to interpret and produce appropriate and meaningful utterances in a specific situation.

Canale and Swain (1980) further defined the concept of communicative competence by adding strategic competence into the model, and it has later been expanded by Bachman (1990) by embracing actual communication strategies called pragmatic competence into the model.
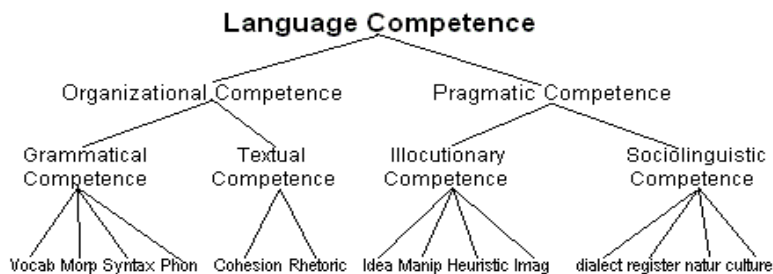
Figure 1: Components of Language Competence (Bachman, 1990, p. 87)

After the new concept of language competence has been introduced, the focus of language teaching has shifted to social and cultural competences in actual context of language use. This is why Bachman encouraged language teachers to include contexts in their language tests because learners' language performance could be influenced by their socio-cultural competence which refers to the degree of awareness of the social convention, norms, and beliefs of the target community.

General Language Proficiency

In 1904, Spearman proposed some models for human intelligence. In his work, he argued that there was one general factor that underlied all cognitive test scores. This is what he called the *g* factor or general intelligence. However, his general factory theory received only little attention due to the complexity of his statistical analysis.

Later on, Prof. John Oller became interested in the implication of Spearman's theory. Oller tested Spearman's hypothesis by using the factor analysis method when analyzing the language construct of the UCLA English as a Second Language Placement Examination. He found that all the subparts on the test were highly correlated to one single factor, in which he called *the general language proficiency factor*.

Based on the general notion of general language proficiency factor that Oller has developed, each of the four language skills manifested itself under one general language proficiency. Based on a number of his studies, it was found that grammatical knowledge highly correlated with *the general language proficiency factor*.

Therefore, Oller (1976; 1976) claimed that the factor that governed the learners' effective language abilities was grammatical competence. Even in integrative tests such as dictation, his studies showed that the best predictor of the students' scores was the knowledge of grammatical structures since it enabled the learners to process and produce the language accurately at a rapid rate.

Evidence for a General Language Proficiency Factor

As mentioned, Spearman contributed a dominant role to the notion of a general factor of intelligence or 'g' factor, and Oller (1983) put forward his idea by applying it into the concept of the unitary language competence hypothesis. This argument has been strongly supported by numerous research studies.

Some of the evidence came from two studies conducted at Southern Illinois University (Farhardy, 1982; Scholz, Hendricks, Spurling, Johnson & Vandenberg, 1980). The subjects who were the students at the Center for English as a second language were asked to complete a language test that contained different sub-sections. Their research suggested that even though there was some unique variance left in the model, the g factor or grammar has explained most of the variance. In sum, their studies have proven the existence of the g factor.

Even though John Oller explicitly argued that only a unitary factor could predict the learners' performance because it accounted for a large portion of the variance in the results of many language tests, he encouraged further research to use a full array of statistical procedures to determine the degree of inter-relationship of different language components.

Divisible Competence Hypothesis

The alternative view of language proficiency is known as the divisible competence hypothesis. Contrast sharply with the unitary hypothesis, there are different constructs underlying language proficiency, and each can be taught or tested separately (Hosley & Meredith, 1979; Powers, 1982).

The hypothesis of divisible competence is largely influenced by a study of Bachman and Palmer in 1982. Bachman and Palmer used a statistical method called MTMM to validate the underlying constructs of the combination of four different tests: a writing test, ten multiple-choice grammar items, an oral interview, and a self-rating method. It has been shown that the loadings of the observed variables on the latent traits largely accounted for two distinct variables which are grammatical and sociolinguistic competence, as seen in their proposed model for language competence.

Another study that shows evidence to support the divisible competence hypothesis was conducted by Fouly, Bachman, and Cziko on three-subtests of TOEFL and Illinois English Placement Test Battery in 1990. Their study revealed the existence of separate language components: oral-aural abilities, structure-reading abilities, and discourse competence.

Nowadays, numerous foreign language tests have been developed based on this model, and their titles tend to isolate language components under four skills: speaking, listening, reading and writing. Under a particular title, however, test users can be misled to the belief that the test construct is made up only of one unique skill.

Alternative View of Language Proficiency

There is still an endless debate about unitary and divisible theories among researchers. While some of them have given up, many of them have come up with an alternative view of language dimensionality. The issue is now more closely related to the problem of how many separate tests are needed to assess general language proficiency and whether productive and receptive skills are equally loaded in the model.

A study that is very influential was conducted by Sawaki, Stricker, and Oranje (2009). In their study, they proposed a model called *the hierarchical model* to explain the language components of the TOEFL-iBT. Based on their data, the model that included a general factor in the highest order and four distinct first-order factors was the best fit model because it accounted for most

variance in the data. To explain a part of the phenomenon, they mentioned the format of the new version of the TOEFL test that includes integrated tasks. Since it requires the examinees to integrate different language modalities when completing the tasks, four distinct language variables -- listening, speaking, reading, and writing -- loaded on a larger factor for the overall language ability.

In sum, there have been three major models of language dimensionality proposed so far -- the unitary trait, divisible language competence, and hierarchical models. However, there is still no consensus among researchers about the model that can best fit the the characteristics of language proficiency.

One problem might have something to do with construct validity because different researchers still do not have general acceptance of what skills can best represent the true picture of language competence. Another issue has arisen from the methods that assess learners' proficiency. Do the measurement tools allow the examinees to have ample opportunities to demonstrate a full range of language abilities through a certain language task, or has their performance been limited to whether they know the correct answers or not? These issues can lead to the discussion in the next section of the most two common methods that have been used to measure the language success of the learners. Knowing the differences between these methods helps explain the different elements of the language that a test claims to measure.

Discrete-Point and Integrative Tests

Over fifty years, language testing theories and practices have followed the changing trends of language teaching methodology. In the 1950s, it was the period when behaviorism was the dominant paradigms, but after the 1970s, the communicative approach brought along a more integrative approach to language teaching and testing. Nonetheless, the two major approaches in language testing that still prevail today are discrete-point and integrative formats.

Discrete-Point Tests

What are Discrete-Point Tests?

The Discrete-point test refers to a test that focuses on one point of a language at a time e.g., segmental phonemes, grammatical structures, or lexis. The discrete-point tests are constructed based on assumptions that language consists of several distinct component parts and those parts can be tested in isolation (Farhady, 1983; Flowerdew & Miller, 2005; Oller, 1979). Consider the examples below.

Item 1: Choose the words with the /p/ sound in them:
 Pen Ben
 Ball Paul
 Bat Pat

(Flowerdew & Miller, 2005, p. 199)

Item 2: Circle the choice that best fits the sentence
 The **boy/boys** strikes the car and runs.
 (Lado, 1961, p. 161)

Item 3: The opposite of **strong** is _____.
 A. short B. poor
 C. weak D. good

(Lado, 1961, p. 189)

Item 1 is an example of discrete-point item whose purpose is to assess whether the learners know how to distinguish between /b/ and /p/. Item 2 is measuring a discrete element of English grammar. All that is required is an understanding of the subject-verb agreement. It does not matter whether the students know the meaning of the word *strike* or not. Item 3 is testing vocabulary knowledge of the learners whether they know the lexical meanings of the words used in the item.

As seen, discrete-point tests are designed to measure only one language element at a time. After this approach had been introduced, many scholars who believed that "The primary function of language is communication" have argued against it (Farhady, 1983) because these pure discrete-point items could not reflect actual language performance.

Controversies and Potential of Discrete-Point Tests

Discrete-point tests have been around for a long time and the example of wide-spread use of such tests are multiple-choice items. Lado (1961) noted that "Language is built on sounds, intonation, stress, morphemes, words, and arrangements of words having meanings. Each of these elements of language constitutes a variable that we want to test" (p. 25). However, it is the second portion of Lado's statement that has become the criticism of those who do not support the use of discrete-point tests. Those people include Lyle Bachman and Adrian Palmer who have coined the term 'target language use' (TLU) and argued that the closer a test reflects TLU, the better it can predict the performance of the learners in the real world.

Even though the controversy surrounding discrete-point tests still prevails, it is argued that there is still a lot of potential in current language classrooms. As Oller (1979) and Oller and Conrad (1971) pointed out, the primary advantage was that this kind of tests was a great way for the teachers to gauge their understanding of what strengths and weaknesses of their students are. For example, a multiple-choice item requires the use of a third-person singular verb in the present tense. If a student misses this item, it is likely that the student does not have sufficient knowledge in the use of English subject-verb agreement. However, one major issue of such test is that the test cannot tell how much knowledge in subject-verb agreement that this student is required to have in order to be able to function effectively in communication. The second advantage of discrete-point tests is that they are very subjective; accordingly, they can yield quantifiable data. They have high scoring reliability, and the scoring process is practical and efficient because they allow a wide coverage of content and items within a limited amount of time.

Integrative Tests
What are Integrative Tests?

The concept of integrative testing emerged to replace a discrete-point test. As its name suggests, integrative test items

require the learners to combine many language elements together for the completion of a task (Oller, 1979).

What are Integrative Tests Measuring?

It was believed that integrative tests could measure the actual aspects of language activities that learners must perform in real life situations (Buck, 2001; Douglas 1989; Jones & Spolsky, 1975; Oller, 1983). By using the integrative approach, the language processing becomes the main focus of assessments because the learners have to be able to understand, process, and produce the language simultaneously (Flowerdew & Miller, 2005). Furthermore, Cooper (1968) emphasized the necessity of incorporating sociolinguistic rules and strategic competence in a test in order to tap the complexity of communicative competence because the learners should know how to use the language appropriately.

Examples of Integrative Tests

A survey of literature shows that traditional testing techniques of the integrative approach include cloze tasks and dictation (Farhady, 1983; Flowerdew & Miller, 2005). These two kinds of tests are seen as integrative because the test takers are required to synthesize the speech by recognizing the phonological, syntactic, and semantic components of the language when completing the tasks. Apart from these two formats, some forms of integrative tests such as oral interviews and writing tests could be considered communicative assessments.

However, the main criticism of the integrative approach is that a number of different test items are required for learners to demonstrate a full range of language abilities, and the results of the test scores might be difficult to be generalized to a different test context. Moreover, some integrative tests tend to involve a large amount of scoring time and rigorous rater training; accordingly, they might not be practical and cost-effective (Farhady, 1983).

Some Considerations of Statistical Analyses for the Concepts of Dimensionality

Now, as it comes to the point where the distinction between discrete-point and integrative assessments has been made, I would like to make the connection between the different uses of tasks and the interpretations of language proficiency. Since the results from factor analyses do not always provide similar results, and different researchers tend to use different methods of factor analytic techniques when analyzing the data, I would like to point out some characteristics of those studies that might help explain why different researchers do not have conclusive ideas about language dimensionality.

First, even though a number of concepts of language proficiency have been proposed, these concepts are still perplexed with what actually constitutes language proficiency. Many linguists have still expressed concern about the lack of one common ground in the theoretical foundation. Accordingly, it is possible that the test items selected to satisfy the same language construct can be highly diverse in content focus. In this regard, the statistical results for the identification of constructs underlying test performance are likely to be different. For example, it is oftentimes said that a test on listening should test only students' listening skills based on what they hear and not to test the learners' vocabulary knowledge or grammatical competence, while some argued that a listening test should be integrated with speaking skills in order to reflect real-life competence, and therefore, with this belief, only assessing listening and speaking skills in integration can measure true communicative abilities.

Second, I think that the controversy on language dimensionality is markedly related to the item formats. As we know, discrete-point tests only test one language skill, while integrative tests measure different language skills at the same time. However, some discrete-test items can actually measure integrative abilities. Consider the two items from the listening section of Michigan Placement Test below. Even though the test developers claim that the objective of this section is to mainly measure the listening ability

in accurately receiving messages with minimal requirement of lexical knowledge, their listening test actually involves a broad range of linguistic knowledge, including the ability in recognizing the sound system of English, knowledge of the grammatical structures in order to be able to interpret the meanings in the oral form, and discourse aspect of the language in order to understand the questions and then choose the most appropriate responses.

> Item 4: If Harvard University accepted him, would he go there or to his state college?
> > a. Yes, he would.
> > b. He'd go to Harvard.
> > c. He went to the state college.
>
> Item 5: Jane could have seen more if she had had a car while she was in California.
> > a. She had a car.
> > b. She will have a car.
> > c. She didn't have a car.

Both items 4 and 5 are measuring multidimensional language skills rather than just the ability to hear and recognize each English word alone. The test takers first need to know the meaning of each word and be able to recognize that the speakers have presented an action that is hypothetical. Then they are required to figure out a consequence of the imagined action or state by choosing the correct structure of hypothetical consequence.

As Henning (1992) has mentioned, different assumptions about language proficiency can lead to different results of the studies. This is a reason why some researchers found multidimensionality of language tests, while others found distinct uncorrelated elements of language competence.

Finally, like reliability, the test of dimensionality is sample dependent (Henning, 1992). Selecting different groups of learners can lead to different conclusions that support different hypotheses (Powers, 1982). Since the nature of a factor analysis is to search for shared variations to reduce the number of unobserved latent variables in a dataset, the information gained from different groups of participants can lead to different factor loading paths. For example, if a group of participants who receive very low scores on a test is selected, residual variance after extracting a general factor

should be very low because there would be no clear grouping among the participants, and therefore the data would support the unitary competence hypothesis. On the other hand, if a highly heterogeneous group of students is selected, it is more likely that the results would indicate divisible competencies due to different dimensions of variances that later can indicate differentiated abilities of the participants in the data. For example, in 1981, Hughes and Woods conducted a study on the Cambridge Proficiency Examination. The test was given to learners from four different L1 backgrounds. The results have revealed distinct language components. Based on this study, I believe that researchers should include clear descriptions of participants' language levels in their study when justifying their conclusions.

In summary, I have pointed out three reasons that might help explain why the results of previous studies are not tenable. The unitary competence hypothesis has never been proven correct, and no language tests can adequately be explained by the divisible hypotheses. I believe that the researchers should make every effort to include detailed description of their research methodology, including the construct(s) of the tests, methods of analyses, and characteristics of the participants. This will give their readers an insight about how their studies are conducted and how they arrive at conclusions that support either the unitary or divisible trait hypotheses.

Conclusions and Implications for Pedagogical Practices

Even though the distinctions between discrete-point and integrative or communicative assessments appear to be too simplistic and theoretical because they can be overlapped in actual practices, I think we can generally agree that communicative testing is preferred in most current language classrooms. However, it is not easy to construct and score a test that can measure authentic language production owing to different contextual constraints such as time, budget, and human resources. These could be the reasons why the exams that assess communicative competence may be rarely used, especially in large language classes.

In fact, there have been many researchers who conducted studies to prove that discrete-point testing could actually be practical and integrative if they are carefully designed because their tests were highly correlated with communicative tests. For example, a study of Ajideh and Esfandiari (2009) was conducted to explore the relationship between multiple-choice vocabulary test and some integrative tests in the Iranian context. Findings of their study showed that the correlation between the two tests were almost .60 which is considered relatively high for the two test formats which were seemingly very different.

In short, what I am trying to argue here is that multiple-choice tests are not always considered discrete-point items. If an item requires an integration of linguistic knowledge with meaning, which simulates the process of natural language use, it should be considered an integrative item. Even though there is still no consensus among researchers that to what extent the boundary between discrete and integrative testing should be drawn. Therefore, I believe that teachers should be encouraged to design test items that can involve different degrees of language skill integration. Consider the following examples.

> Item 6: Unless you .... the project in time, you are not going to pass this course.
> > a. submit
> > b. do not submit
> > c. submitted
> > d. did not submit

As seen, the students have to read through the complicated language structures and understand how different language elements are used and interpreted in the sentence before they can choose the correct answer; therefore, the teachers may find that this item is testing different language skills at the same time.

> Item 7: Did you read that article about our company in "Newsweek"?
> > a. No, not yet. What is it about?
> > b. I agree that this is a very busy week.
> > c. I'm not sure. I couldn't read his mind.
> > d. Great. I'm looking forward to receiving it.

This listening item measures the students' ability to understand spoken English in a real-world workplace situation. Even though it is true that this listening item might not be able to test the learners' actual communicative ability, it can, at least, measure the learners' comprehension of the language used in a given context. In order to choose the correct response, the students have to integrate the communication skills of phonology, vocabulary, semantics, syntax, and pragmatic.

Nonetheless, this is still an overarching problem of how to validly and reliably assess authentic language production. Even though language teachers and researchers agree that communicative performance is a multidimensional construct, the research implications of the construct's multidimensionality are still not well-understood. However, this does not mean that researchers should stop measuring communicative competence or conducting the studies on language dimensionality. What we should do is to use a synthesis of previous attempts to describe the dimensions of language competence and the analysis of the language proficiency notions to build a conceptual model that can fit our context and, at the same time, be practical, and economical. Assessment of communicative skills could require a detailed rubric and rigorous training to accurately observe the performance of the learners in a communicative context.

Finally, what I try to present here is that we cannot completely accept or reject any theories, hypotheses, or any forms of language tests. Instead of trying to propose one single well-accepted approach for teaching language and testing students' language proficiency, we can embrace different statistical results and try to find directions for future research that can advance understanding and conceptualizing language performance and how to measure it. In this way, we can go beyond teaching language per se and extend our perspectives, so that we can help second language learners develop themselves. This would indicate the need for the development of appropriate teaching strategies and the implementation of appropriate syllabi.

The Author

Suthathip Thirakunkovit, Ph.D. graduated from Purdue University in 2016. She is currently a lecturer in the English and Applied Linguistics Programs at the Faculty of Liberal Arts, Mahidol University, Thailand. Her research interests broadly focus on test validation, corpus linguistics, and second language writing.

References

Ajideh, P., & Esfandiari, R. (2009). A close look at the relationship between multiple choice vocabulary test and integrative cloze test of lexical words in Iranian Context. *English Language Teaching, 2*(3), 163-170.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

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

Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign language students. *Testing the English Proficiency of Foreign Students*, 36.

Cooper, R. L. (1968). An elaborated language testing model. *Language Learning, 18*(3), 57-72.

Douglas, D. (1989). Testing listening comprehension in the context of the ACTFL proficiency Guidelines. *Applied Language Learning, 1*(1), 53-73.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly, 16*(1), 43-59.

Farhady, H. (1983). On the plausibility of the unitary language proficiency factor. In J.W. Oller, Jr. (Ed.). *Issues in Language Testing Research.* Massachusetts: Newbury Publishers.

Flowerdew, J., & Miller, L. (2005). *Second Language Listening: Theory and Practice.* Cambridge: Cambridge University Press.

Fouly, K. A., Bachman, L., & Cziko, G. (1990). The divisibility of language competence: A confirmatory approach. *Language Learning, 40*(1), 1-21.

Henning, G. (1992). Dimensionality and construct validity of language tests. *Language Testing, 9*(1), 1-11.

Hosley, D, & Meredith, K. (1979). Inter- and intra-test correlates of the TOEFL. *TESOL Quarterly, 13*(2), 209-217.

Hughes, A., & Woods, A. (1981). Interpreting the performance on the Cambridge Proficiency Examination of students of different language backgrounds. In A. Hughes & D. Porter.

Current Developments in Language Testing. London: Academic Press.

Hymes, D. (1972). On communicative competence. *Sociolinguistics,* 269-293.

Jones, R. K., & Spolsky, B. (1975). Testing Language Proficiency. ERIC.

Lado, R. (1961). Language Testing: The Construction and Use of Foreign Language Tests. New York: McGraw-Hill Book Company.

Oller, J.W., Jr. & Conrad, C. (1971). The cloze technique and ESL proficiency. *Language Learning, 21*(2), 183-196.

Oller, J. W., Jr. (1976). Cloze, discourse, and approximations to English. In K. Burt & H. C.

Dulay (Eds.). *New Directions in Second Language Learning, Teaching and Bilingual education.* TESOL: Washington, D.C.

Oller, J.W., Jr. (1976). Evidence for a general language proficiency factor. *Die neueren Sprachen, 75,* 165-174.

Oller, J.W., Jr. (1979). *Language Tests at School: A Pragmatic Approach.* London: Longman.

Oller, J.W., Jr. (Ed.). (1983). *Issues in Language Testing Research.* Massachusetts: Newbury Publishers.

Oller, J.W., Jr. (1983). "G", "What is it? In A. Hughes & D. Porter. *Current Developments in Language Testing.* London: Academic Press.

Powers, D. (1982). Electing samples for testing the hypothesis of divisible versus unitary competence in language proficiency. *Language Learning, 32*(2), 331–335.

Rea, P. (1985). Language testing and the communicative language teaching curriculum. In Lee, Y. P. et al. *New Directions in Language Testing.* Oxford: Institute of English.

Sawaki, Y., Stricker, L., & Oranje, H. (2009) Factor structure of the TOEFL internet-based test. *Language Testing, 26*(1), 5-30.

Scholz, G., Hendricks, D., Spurling, R., Johnson, M., & Vandenberg, L. (1980). Is language ability divisible or unitary?: A factor analysis of 22 English language proficiency tests. *Research in Language Testing.* Massachusetts: Newbury House.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292.

Spolsky, B. (1981). Fourteen years on-later thoughts in overall language proficiency. In A. Hughes & D. Porter. *Current Developments in Language Testing.* London: Academic Press.

Vollmer, H. J., & Sang, F. (1983). Competing hypotheses about second language ability: A plea for caution. In J. W. Oller, Jr. (Ed.). *Issues in Language Testing Research.* Massachusetts: Newbury Publishers.