Applied Language Learning

VOLUME 9 · NUMBERS 1&2
1998

APPI LEARN 1998 VOLUME 9 U

Applied Language Learning

Professional Bulletin 65-98-1 & 2

1998 **Volume 9 & 2** **Numbers 1**

Lidia Woytak
Editor

Defense Language Institute Foreign Language Center

Presidio of Monterey, CA 93944-5006

In Memory of

Richard A. Woytak

Who always has believed that

Applied Language Learning

Will Create a Better World for all of Us

Applied Language Learning

PB 65-98-1 & 2

The mission of Professional Bulletin 65-98, *Applied Language Learning (US* ISSN 1041-6791), is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences.

Applied Language Learning, published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, presents professional information. The views expressed herein are those of the authors, not the Department of Defense or its elements. The content does not necessarily reflect the official US Army position and does not change or supersede any information in official US Army publications. *Applied Language Learning* reserves the right to edit material.

By Order of the Secretary of the Army:

DENNIS J. REIMER
General, United States Army
Chief of Staff

JOEL B. HUDSON
Administrative Assistant to the
Secretary of the Army
05596

DANIEL D. DEVLIN
Colonel, United States Army
Commander
DLIFLC and POM

Further reproduction is not advisable. Whenever copyrighted materials are reproduced in this publication, copyright release has ordinarily been obtained only for use in this specific issue. Requests for reprints should be directed to the authors.

Availability

To access *Applied Language Learning* on the Internet type:

http://lingnet.army.mil

Additionally, you may obtain the journal on microfilm from ERIC Clearinghouse on Language and Linguistics, Center for Applied Linguistics, 1118 22nd Street, NW, Washington, DC 20037.

Bulk rate postage is paid at DLIFLC. The basis of official distribution is one copy per training instructor and one per five military linguists.

Postmaster

Send change-of-address information to:

Applied Language Learning
Defense Language Institute Foreign Language Center
Presidio of Monterey, CA 93944-5006

United Parcel Customers

Location is *Applied Language Learning*
Tin Barn
Bldg. 518, Room 7
Presidio of Monterey, CA 93944-5006

Readers

Contact Editor, Dr. Woytak, by:
E-mail: woytakl@pom-emh1.army.mil
Telephone: (831) 242-5638
DSN: 878-5638
Fax: (831) 242-5850

Wordprocessing and graphics *Specialist Richard S. Slone*
Private Timothy S. Hughes

Cover design *Barney Inada*

From the Editor

Reviewers for *Applied Language Learning*

The individuals listed below served as reviewers of manuscripts submitted to *Applied Language Learning* in 1998. We express our gratitude for expert service to:

Martha S. Bean
San José State University
Christine M. Campbell
Defense Language Institute
Foreign Language Center
John B. Carroll
University of North Carolina
Marianne Celce-Murcia
University of California
Los Angeles
Dan Douglas
Iowa State University
Roderic A. Gale
Defense Language Institute
Foreign Language Center
Robert C. Gardner
University of Western Ontario
Rafael Gomez
Monterey Institute of International Studies
John S. Hedgcock
Monterey Institute of International Studies
Evelyn Hatch
University of California, Los Angeles
Eli Hinkel
Seattle University
Gordon Jackson
Defense Language Institute
Foreign Language Center
Renee Jourdenais
Monterey Institute of International Studies
Dorry M. Kenyon
Center of Applied Linguistics
Stephen D. Krashen
University of Southern California
Masaki Kobayashi
University of British Coumbia
James F. Lee
University of Indiana
Virginia LoCastro
International Christian University
Barbara Matthies
Iowa State University
Naoko Matsuo
Monterey Institute of International Studies
Denise Murray
St Jose State University
Paul Nation
Victoria University of Wellington
Rebecca Oxford
University of Alabama
Richard Schmidt
University of Hawaii-Manoa
David J. Shook
Georgia Institute of Technology
Roberta J. Vann
Iowa State University
Leo Van Lier
Monterey Institute of International Studies
Dolly J. Young
University of Tennessee

Applied Language Learning

1998 **Volume 9** **Numbers 1 & 2**

Articles

1 Language Aptitude Testing:
Learners and Applications
James R. Child

11 Zero-Based Language Aptitude Test Design:
Where's the Focus for the Test?
Pardee Lowe, Jr.

31 The Modern Language Aptitude Test for Predicting Learning
Success and Advising Students
Madeline Ehrman

71 Factors in the Prediction of Achievement and Proficiency
in a Foreign Language
Richard Sparks, James Javorsky, Jon Patton, and Leonore Ganschow

107 Essay Scores as Instruments for Placement and
Advancement in an Intensive English Program
Lynne Davis, Ruth Johnson, and Floyd Olive

News and Views

121 Interpreter in Action: Interview with Lieutenant Colonel
Richard Francona (Retired)

Reviews

147 Cots: Teaching by Chatting.....Leo Van Lier
149 Kenny and Savage (Eds.): Language and Development:
Teachers in a Changing World.....Kevin W. K. Chu

151 Lee and VanPatten: Making Communicative Language Teaching Happen: Directions for Language Learning and Teaching.....Phillip A. White
154 Gates: The Road Ahead.....Roderic A. Gale

General Information

157 Calendar of Events
163 Information for Contributors

Applied Language Learning
1998, Vol. 9 Nos. 1 & 2, pp. 1-10

Language Aptitude Testing

Learners and Applications

James R. Child
National Cryptologic School

The present article reviews language aptitude testing dilemmas both from the perspective of the National Security Agency (NSA) and of all the government organizations with language missions. First, it considers the possible need to cross-train linguists in government employ, sometimes from "difficult" into "easier" languages, but more frequently in the other direction. In so doing, it recommends which of three available aptitude measures is most suitable. Second, it takes up the relationship between language aptitude and the Interagency Language Roundtable (ILR) skill levels: Which test is appropriate for what skill? Can an aptitude test by itself or together with other measures predict learning success at Level 3 and up? Finally, the article treats the varying degrees of distance between English and other (selected) languages with reference to the aptitude model best suited to the purpose.

The need has become increasingly acute in U.S. society for persons with the four-skills described in the Interagency Language Roundtable (ILR) statements: Speaking (S), Listening (L), Reading (R) and Writing (W). It would seem arguable that any society would be enriched when a significant number of its members can attain at least a Level 2 competence in a given language in one or more skills rated over a range of six levels. Whether the enrichment is a matter of extending personal horizons for social and cultural purposes, or developing skills for the workplace, is less immediately important than the learning attainment per se.

Candidates for Aptitude Testing

The point of departure in considering NSA's requirements and responses to them is the assumption that the language workforce is made up by and large of native speakers of English who have formally learned or otherwise acquired one or more foreign languages. Naturally those persons already skilled in languages the agency requires will likely spend significant portions of their careers processing texts in those languages, at the same time attempting to improve their English writing skills, in order to present better what they have understood, in translation, summary or other form.

However, as changing conditions dictate, it does happen on occasion that linguists in the hiring pipeline or already employed need to be cross-trained to other languages. This problem, fairly widespread in the "world of work" in general, regularly surfaces at NSA. Consideration is given below to the dimensions of the problem and the ways in which it can be handled effectively.

On-Board Working Linguists

Persons who have been hired as linguists and spent sufficient time doing solid professional work in second (or, less frequently, second and third) languages have shown themselves to be excellent candidates for the learning of other, usually "exotic" tongues. The kinds of language processing typical of the "world of work" at many government agencies include translation or interpretation as "top-down" skills, with subsidiary requirements to extract from (whole) products of those skills the critical information needed, in the form of summaries, paraphrases and the like. These activities may be viewed in the government frame of reference as performance measures. That is to say, they are carried out according to the letter and spirit of mission requirements which by their nature demand familiarity and experience with a wide array of content areas (or, "real-world knowledge"). Linguists with a "track record" in coping with one or several of these areas in language X have shown that they can bring much of that knowledge and experience to language Y, with the caveat that Y share a fair amount of basic structural features with X. Absent those commonalities, no amount of subject matter familiarity can in itself predict success in mastering a difficult language. Therefore NSA has opted to administer two aptitude tests—one, the Defense Language Aptitude Battery (DLAB) developed by Petersen and Al-Haik in 1973 as a measure to indicate likelihood of success in learning languages "structurally close" to English; the second, a test James Child developed at

NSA in 1973 called VORD (not an acronym) as an instrument for placing persons in classes in which the language is vastly different in structure from English. The notion of "distance" between and among languages in terms of difficulty will be discussed at a later point; for the moment it is enough that the two tests are administered to employees in conjunction with other measures relevant to cross-training, and that (at least in the case of DLAB) aptitude testing may simply be supportive of other indicators.

Prospective Linguists

Persons in this category normally have one thing in common: little experience in doing the kinds of language work many agencies require. Thus they must spend considerable time in becoming familiar with various specialized topic domains even when they would be using the languages which they bring with them. Without the transferable skills of on-board persons, aptitude testing might prove a critical element in helping managers and senior level technical experts make sound cross-training decisions in placing new hires in language training for which they were not originally programmed.

Language Uses: Skills and Levels

The three skills in demand in agencies taken as a whole are speaking, reading, and listening. (A fourth, writing, has always been somewhat marginal for the government). They are covered for government purposes, in the ILR definitions, in a six-tier system, from Level 0 (extremely limited memorized skills) to Level 5 (that of a highly educated native), with "plus-levels" in between each base level.

Of the skills, reading and listening (the receptive skills) are critical for those agencies which do not have major missions requiring interaction with speakers of (particular) languages, while speaking must obviously be added when such interaction is called for. As to the levels of attainment most frequently in demand in the world of work, it is fair to say that Levels 2 and 3 come into play most often in day-to-day operations, with a Level 4 competence occasionally required.

What does all of this imply for language aptitude testing? First, Level 2 skills should be targeted first, for Level 1 language is generally reduced in content and form to phrase- and sentence-length units containing material usually of little intrinsic interest to serious study: greetings; weather information; arrivals and departures of carriers and the like; all amenable to brute memory. Level 2, on the other

hand, is equally concerned with the transmission of facts, but facts embedded in formal systems of grammar and lexicon requiring much more than mere memorization. News reports or domestic and international events; instructions on how to do or make something; detailed directions for getting to a distant place are excellent examples both of realized Level 2 texts and of the skills required to process them. Thus, well conceived aptitude measures will be designed against the demands of second-language texts of these kinds in any or all skills. The models in current use the Modern Language Aptitude Test (MLAT) developed in 1959 and the DLAB and VORD mentioned earlier work well as predictors of success in learning languages up to Level 2 in the reception mode—reading skills for VORD and most of MLAT and listening skills (among others), for DLAB. (Of the three, MLAT and DLAB were validated decades ago, while VORD, promising in several respects, is still undergoing validation at a government agency). VORD does appear to have an advantage in that persons doing well on this test generally succeed in mastering the syntactic patterns of languages vastly different in structure from English. However, VORD, as noted, is tailored to predicting success in reading only, and in languages employing the Latin alphabet.

The first question to be taken up in the sections below is whether VORD (or any aptitude measure) can forecast attainments beyond Level 2 in any language, and if it can, to which level.

Aptitude Tests as Predictors Beyond Level 2

Existing aptitude measures—singly or severally— have proved to be reasonably satisfactory predictors of success at Level 2 for the three skills in question. But does that suggest a comparable outcome at Level 3 or higher?

There does not seem to be a great deal in the literature of language aptitude testing specifically bearing on success past Level 2, although the desirability of higher levels of attainment is obvious, especially in regard to performance in the work place. The difficulty, though, is to include in the test design those language elements characterizing texts at Level 3 (and higher) as stated or implied in the ILR descriptions. These include references to "...hypothesis, argumentation and supported opinions," the language of which is likely to be relatively rich in lexicon and culturally sensitive. Current aptitude models do not reflect these features; in fact, it is difficult to see how they could be built into tests in the (relatively) short time provided for their administration. It would seem that there are only a few alterna-

tives for devising such instruments: a battery including a current aptitude model accompanied by a measure or measures to elicit cultural sensitivity or other psychological aspects; an extended "pure" aptitude test in which, say, Level 3 tasks are embodied in a sophisticated syntax and lexicon, the mastery of which would be exceedingly time-consuming, quite possibly to the point of impracticality.

Prediction of higher level success is but one of the goals of language aptitude testing. Another is the design of a test or test battery which indicates whether the channel of communication (i.e., through the eye or the ear) makes an essential difference. There is no doubt that individuals have preferences in this regard and that memory retention may be a problem for the listening channel, especially if passage replay is not permitted. However, since the item structures in two of the measures are confined to very short language segments at the clause level and below (MLAT and DLAB) the memory load may not seriously affect test performance. VORD on the other hand does include longer texts, especially the ones with the planned blanks (CLOZE-like texts), but, as already observed, it was not designed to test listening comprehension at any point. (For a study of the relationship between MLAT and VORD, see Parry & Child, 1990).

Interestingly, though, VORD at least seems to have some "cross-over" predictive value. Preliminary analyses have been carried out by two government agencies on the performance of language learners on VORD vis-a-vis their subsequent levels of attainment in multi-skill language courses. On the surface, such a result may seem improbable, since speech has a tempo, pattern of pitch or stress, and on occasion tonality, poorly replicated if at all in the writing system (which of course has its own peculiarities often unrepresented in speech). However, language is ultimately a question of communication (very possibly limited to Level 2 for present purposes), so that expectancy based on knowledge of subject matter or familiarity with a situation (sometimes referred to as "semantic feedback") can override problems created by a difference of channel. Thus, VORD, and the other aptitude models as well, may have intrinsic features which allow for cross-channel inferencing. Much more investigation will be required to determine the validity of this hypothesis.

Distances Between and Among Languages

The entire "language aptitude" enterprise could falter in the absence of a comprehensive overview of similarities and differences

among the major languages of the world. There have been over the years a number of attempts to categorize languages in terms of their presumed difficulty; which is to say, how hard they are to learn for native speakers of English. Several of these efforts have in fact been officially blessed within a number of government agencies because they have a certain face validity and have proved useful as general guidelines. However, they do not specify what features of which languages can be expected to cause trouble for learners and which are similar to, or not very different from comparable English features. To lend greater precision to a "global assessment" system it is necessary to determine which major linguistic features of the so-called "hard" languages make learning problematic for English speakers and which lend themselves to (relatively) easy transfer into English.

The following paragraphs set forth what are generally agreed to be the major components of language (whether in speech or writing); the matrix at the table may be useful in following the explanations: (A) phonology (with provision made for written representation); (B) grammatical system, covering what have been traditionally called morphology and syntax; and (C) semantics, taking in meaning in all of its textual representations. These three phenomena are then ordered in such a way as to indicate relative distances of foreign languages from English: Near (1), Middle (2), and Remote (3).

How does all of this relate to language aptitude? The answer is, in a fundamental way, that learning difficulty is tied to the degree in which the object of learning resembles something already known. In the present case it is not enough to say that language X differs greatly from English without specifying the nature of those differences. Let us consider some examples.

A need arises to train a number of individuals in German, a language historically related to English. These persons have had either Spanish or French for one or two high school semesters, not enough to give them sufficient language-learning experience to get off to a fast start in German. That language offers sufficient difficulty to warrant aptitude testing, but is not so daunting overall as to demand very high linguistic skills. Specifically, German uses the Roman alphabet and does so in a way that is roughly isomorphic with the spoken language, hence it is a "near" relative to English (Distance 1). The grammatical system, while sharing many features with English, is sufficiently different syntactically to warrant a "middle"

Table

*Elements of Textual Distance from English to Language to be Acquired**

Textual Elements	**Distances**		
	Near (1)	Middle (2)	Remote (3)
Writing System as representation of spoken language (A)	◆Language in Roman alphabet, with reasonably close fit between "letters" and "sounds." Greek and Cyrillic alphabets may also be included, as they also have close relationships with the associated spoken forms.	◆Letters in Roman alphabet at some removal from "sounds" to be repre sented. ◆Letters in non-Roman alphabet subject to considerable variation in shape, according to positions in words, but generally calibrated to sound system.	◆Non-Roman letters change drastically both in terms of positionin words and in clusters they enter into. ◆Syllabary of considerable size. ◆Character set, without correspon-dence with sound system.
Grammatical system as frame-work for commu-nication (B)	Surface syntax and vocabulary contain more features similar than dissimilar vis-a-vis English. "True friends" outnumber false ones and there are relatively few features that cause great difficulties for English speakers. Unit lengths of formal structures (e.g., clause, sentence) not greatly different from those of English.	While many grammatical and lexical features resemble those of English there are many points invoving word order, nominalization and the like which English either lacks or uses in more restricted circumstances than is the case in the stances than is the case in the studied language. Lengths of formal units may vary considerably (e.g., much longer, somewhat shorter) than those of English.	Great numbers of grammatical features totally unfamiliar to English speakers: verb/noun relationships, for example, are not like those of English (or other major West European languages). Clause andsentence structure may be marked by affixes or particles rather than by a system of frequent punctua-tion.
Semantic system as cultural outlook (C)	Considerable cultural overlap between language and English; hence, expectancy of learner not basically misplaced.	Cultural milieu of studied language different in many respects from that of English. Still, culture not totally alien.	Culture vastly different from English speaker's; For example, social distinc-tions for caste, gender and the like are formally present in every kind of text.

* Each language may be described by three sets of alphanumerics, e.g., Chinese might be A3/ B2/ C3.

rating (Distance 2); and, in terms of the semantic system, the language expresses a cultural outlook with much in common with that of (American) English speakers, but enough difference to require, again, a Distance 2. Thus, German may be reasonably characterized in the matrix as A1/B2/C2.

Of the available aptitude instruments, which might be the one of choice in this case? Without going into a detailed comparison of the three—for which there is not enough time in this paper—it may be noted that MLAT and DLAB items are confined to word and phrase segments roughly similar to English in length and part-of-speech category, while VORD has sentence- and paragraph-length items expressed by a syntax truly alien to most language learners. Thus, either MLAT or DLAB are preferable to VORD for German.

Turkish may be taken as an example of a much more difficult language system. It does use the Roman alphabet, but with some additional orthographic devices not found in English. Nonetheless, the alphabet enables close coordination between speech and writing, hence rates the digraph A1 (phonology/written representation relatively "near" in terms of learner difficulty).

The grammatical system is a very different matter. Many constructions which would require in English and most European languages verb tenses—past, present, future, and others—employ nouns derived from verbs. These come as a distinct shock to most American students of the language, especially when they are embedded in sentences of 80 to 90 words, a quite usual occurrence in Turkish. Thus, this system can be properly noted B3 (grammar is "remote" from English and a major source of difficulty).

A similar judgment could also be rendered for the semantic system, even though the grounds for the decision are quite different. Turkish has been greatly enriched by its huge number of lexical borrowings from Arabic and Persian and its incorporation of them into the complex Turkish culture. This very complexity can and does cause difficulty to neophytes especially at Levels 2 (high) and 2+ (toward the end of the spectrum in which aptitude testing is likely to be effective). A rating of C, "remote," may be applied here as well. In short, the overall characterization for Turkish is A1/B3/C3.

A final example might be Japanese. This language depends upon both a syllabary (i.e., a consonant plus vowel representation of speech in the writing system) plus a vast number of characters taken over from Chinese. The latter, in addition to imposing a huge memory burden, are not systematically aligned with spoken language segments, with the result that the phonological component, as such, blends

with the lexical (a situation which also pertains to some extent in Korean). The only possible description is "phonology/writing system remote from English," or A3.

Japanese grammar is formidable as well. The verbal system has two basic tenses, but a great number of forms expressing feelings and attitudes of the speaker. These, too, are mostly alien to English, hence, "remote" to the struggling learner. B3 is an apt characterization for the verbal system as well as a number of other Japanese grammar phenomena.

Finally, Japanese culture differs greatly from American culture as expressed in its use of devices reflecting social status. "Polite" vs. "abrupt" verbal forms reflecting particular kinds of social interaction are essential in communication, therefore "musts" for the learner. Again, "remote" is the best description here for semantic differences based on Japanese culture. Thus, Japanese can be captured via the matrix as A3/B3/C3.

The three examples above (German, Turkish, and Japanese) barely scratch the surface of possibilities. They are easily labeled because the characteristic features are so clear-cut. A number of other languages are significantly harder to label: for example, the Cyrillic alphabet used in Russian and several other Slavic languages is problematic for some learners, not so for others. Should it be labeled as A1 or A2?

Such questions are reminders that the designations are, ultimately, relative to a degree and for present purposes reflective of "other languages" considered in the light of English. Literate native speakers of Russian, for instance, would have no problem with the Cyrillic alphabet as used in certain other Slavic languages (though they may have difficulties when it is employed in non-Slavic languages). Nor would a literate native speaker of Turkish have great difficulty with the syntactic patters of Hungarian or Mongolian which in many aspects resemble those of his own language.

References

Carroll, J. B., & Sapon S. M. (1959). *Modern Language Aptitude Test* (MLAT). New York: The Psychological Corporation.

Child, J. R. (1973). *VORD*. Washington, DC: U.S. Department of Defense.

Parry, T. S., & Child, J. R. (1990). Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In T.S. Parry, & C.W. Stansfield (Eds), *Language aptitude reconsidered* (pp 30-66). Englewood Cliffs, NJ:

Prentice-Hall Regents.
Petersen, C. R., & Al-Haik, A. R. (1977). *The development of the Defense Language Aptitude Battery* (DLAB).

Author

JAMES R. CHILD. Senior Research Linguist, National Cryptologic School, Department of Defense. Specializations: language aptitude research, analysis of textual proficiency and performance levels, and translation theory.

Applied Language Learning
1998, Vol. 9, Nos. 1 & 2, pp. 11-30

Zero-Based Language Aptitude Test Design
Where's the Focus for the Test?[1]

Pardee Lowe, Jr.
National Cryptologic School

Aptitude test design has generally led to a "one-test-fits-all" approach, with tests undistinguished as to whether they predict success in language generally without regard to skill modality or whether they are better for one skill over others.[2] Moreover, aptitude test design has failed to distinguish among various levels of possible attainment so that in predicting success one is unable to say whether someone with a high aptitude will generally attain Interagency Language Roundtable (ILR) level 3 in speaking just because she or he has an aptitude for learning a language; nor can one say in what language or type of language. At least one aptitude test—VORD (not an acronym)[3]—attempts to identify those with an ability to learn languages with more complicated grammatical structures like Russian or Japanese. However, its designer, James R. Child, points out that VORD does not predict well the ability to learn tone languages like Chinese. This article looks at the many questions aptitude test design has failed to answer and queries whether it would be possible to return to ground zero (zero-based test design) and design tests that could address them. The questions raised here have not generally been talked about in the literature, but are routinely asked by managers of U.S. Government language programs; so, while sometimes naive, they are actual and important to the success of the U.S. Government's language training effort.

Previous Tests and Their Components

There are several language aptitude tests: The *Artificial*

Language Aptitude Test (ALAT), The *Defense Language Aptitude Battery* (DLAB), The *Elementary Modern Language Aptitude Test* (EMLAT), The *Modern Language Aptitude Test* (MLAT), The *Pimsleur Language Aptitude Battery*, and *VORD*. For government use, three tests, however, should be removed from any serious consideration at the outset: the EMLAT and the *Pimsleur Language Aptitude Battery* that target a lower age group than that of government workers, and ALAT (derived from the MLAT) which is too Eurocentric in design and was replaced at least at the Defense Language Institute Foreign Language Center (DLIFLC) 25 years ago by the DLAB. In other words, the models in current use are instruments better designed for government needs.

While the results from the three remaining tests, DLAB, MLAT, and VORD, have been useful to a degree, the government would like more accurate predictors. The MLAT as administered at the Foreign Service Institute (FSI) of the Department of State correlates consistently around .50 with end of training scores (reported on the ILR scale), and, therefore, accounts for 25% of the variance in student learning.[4] I believe that one needs to characterize aptitude comprehensively and to build more accurate instruments for identifying the ability to learn and to use the language.

We have listed the tests. What then of the tasks they employ to determine language aptitude at least as conceived at the time of test design? An overview of the components of various aptitude tests reveals a divergence of opinion among aptitude test designers on the factors and tasks which contribute most strongly to the construct "language aptitude" (see Table 1).

There seems to be disagreement about what constitutes the construct "language aptitude" itself. We would amplify the phrase "language aptitude" by the words "ability to learn and to use language." While this definition is workable, it by no means replaces a need to more fully define and characterize the ability generally referred to as "language aptitude." The fact that past aptitude test designers have drawn on rather divergent predictors to determine aptitude for learning another language suggests either that there is no agreement on what constitutes the construct "language aptitude," nor that there exist numerous possible predictors that could serve as components in determining the construct, or even that the focus of past test designs has been clear.

One of the most important tasks of the language aptitude testing discipline is to try to reach agreement on a definition of language aptitude. Such a definition should be broad enough to cover the challenge of learning a new language system per se and the ways that system

Table 1

Aptitude as Previously Defined

Sections	Pimsleur	EMLAT	MLAT	DLAB	VORD
Bio Section				X	
Finding Rhymes	X				
Foreign Language Grammar				X	
Grade Point Average	X				
Hidden Words		X			
Interest	X				
Language Analysis	X				
Matching Words		X			
Nouns in the Artificial Language (AL)					X
Number Learning		X	X		
Paired Associates		X	X		
Phonetic Script			X		
Phrases & Sentences in the AL					X
Recognition of Stress Patterns				X	
Sound Discrimination	X				
Sound Symbol Association	X				
Spelling Cues			X		
Text with fill-in-the-blanks in AL				X	
Verbs in the AL					X
Vocabulary	X				
Words in sentences			X		

is reflected in natural texts. Reaching broad concurrence on the meaning of aptitude would ultimately result in the production of more accurate tests.

Points for Consideration

Since the times of the writing of the aptitude tests listed above, our concept of aptitude has undergone a radical transformation, more specifically, an expansion. Motivation, learning styles and strategies, as well as teaching styles and methods are part and parcel of the language aptitude picture for many researchers and test designers today. Just how these aspects of language aptitude interrelate is unclear, but Figure 1 suggests a possible view which regards language aptitude as having a core (beyond what it has traditionally contained as the questions later

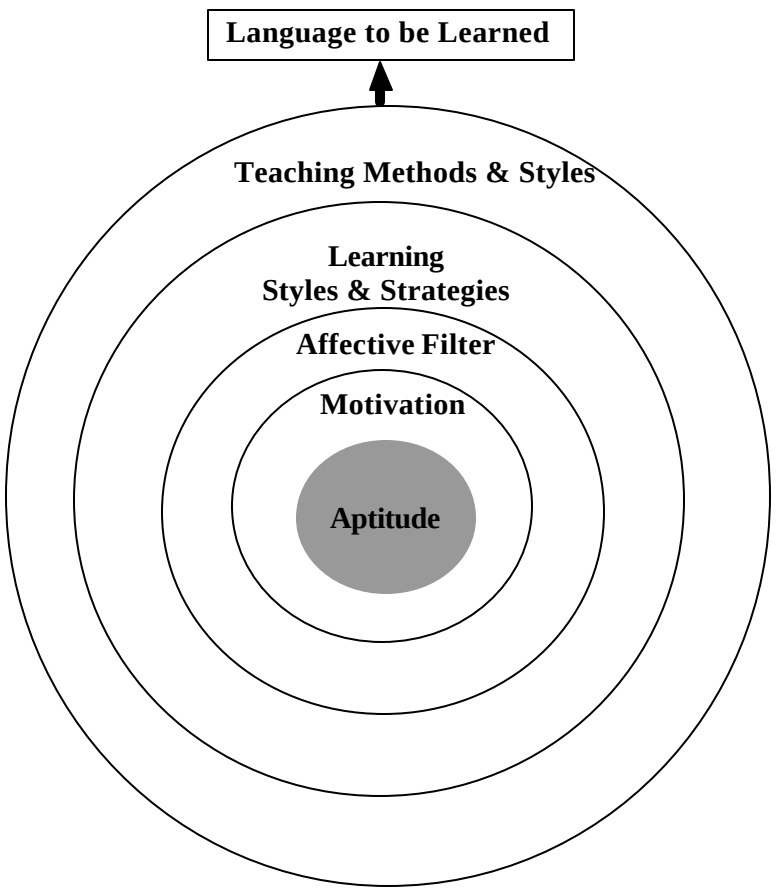

Figure 1. Aptitude and its facilitators

in the article make clear), but which also regards language aptitude as facilitated by other factors, such as motivation, the affective filter and learning styles and strategies on the part of the learner and teaching *methods* and *styles* on the part of the teacher. Their ordering is conceived from the learner's perspective first and the teacher's second. Aptitude resides at the core, but it may be potentiated or attenuated by motivation. We know that high motivation often overcomes low aptitude, while low motivation can defeat high aptitude. On the other hand, a low affective filter allows the full thrust of aptitude and motivation to work, while a high affective filter impedes that combination. Other facilitators depending on the suitability of each to the learner are his learning styles and strategies and the teaching methods and styles of his teacher. Discussion of the interrelationships and contributions of the respective factors leads to an expanded perspective on aptitude.

In this positive ferment, however, there exists a concern that one is failing to focus on some issues pertaining to aptitude test design. This article is concerned with the logical focus of such a design and steps back to regard it from a visual, holistic perspective. Rather than assuming the components employed in aptitude tests so far or the construct that underlies them are adequate, I clear the slate — adopting a zero-based test design. The term *zero-based* is derived from justifying budgets and programs from the ground up; that is, assuming that nothing is sacred and that everything must be re-justified in order to receive funding.

The adoption of the term *zero-based* calls for starting over with new hypotheses. Therefore, I will look at how tests have been designed in the past and ask whether future tests should be designed quite differently, especially in regards to the questions that government managers ask of those who propose to use aptitude tests to identify government employees who could best learn another language. I am not sure that questions government managers pose regarding aptitude test results are always reasonable, nor that one test design can respond to them all. But they form a new point of departure. Moreover, these questions lead to another, perhaps more overarching one:

Q0: What Is to Be the Focus of the Test?

This may be the single most crucial question language aptitude test designers face. Of course, earlier language aptitude tests attempted to define and operationalize each designer's concept of the construct. An approximate .50 correlation between aptitude test scores and exit proficiencies suggests that one might be able to do so more fully.[4]

Moreover, our understanding of the construct has expanded. Finally, the real world poses some rather different questions about language aptitude today than were asked in the past. To answer the question about focus more fully, we list questions an aptitude test could answer in a government context, and we provide background to these questions. These are actual questions government managers have asked when aptitude test results have been used to determine who benefits from training.

A metaphor comes to mind here, that of a runner approaching a course of hurdles. In the future aptitude test we envision here, each question asks if the runner can leap the next hurdle and move on. The test should answer each question as clearly and accurately as possible. These questions doubtless derive from a very broad concept of the construct "language aptitude," one broader than current tests are designed to tap.

Q1: Can an Aptitude Test Tell Us That Someone Can Learn a Foreign Language?[5]

Each year the government devotes significant resources, human and monetary, to training its employees including the military in learning foreign languages. With diminishing resources, the question, "Who is most likely to succeed?" grows more acute.

Q2: Do Languages Have "Personalities," and Is It Possible to Match Language and Person; What is the Effect of Such a Match?

This question, referring to motivation, is not as well addressed in test design as it might be. It asks not just who has a good likelihood of learning a foreign language, but does that person have an interest in that language. Does it match his personality? Does the country, the culture, the ethos of its people, the way they think and act fascinate him? Here the type of motivation plays a role: instrumental, integrative, assimilative? Can a test identify these matches and their strengths?

Q3: How Difficult a Language Can the Examinee Handle?

Early in providing language training, the U.S. Government discovered that Americans learned some languages like Italian and Swedish more readily than they learned German, Russian, or Turkish. In other words, the latter took longer for Americans to master (see Table 2). To deal with this discovery, the government has set aside a longer period for training students to acquire these tongues, allowing

approximately 6 months for the easiest, 9-12 months for somewhat harder ones, and almost 30 months (often including a year in a country) for the hardest. Can one, therefore, predict who will be most successful, not just learning a language, but learning one of a specific category?

Q4: What Language Type(s) Can the Examinee Most Likely Master?

The division of languages into difficulty groups (see Table 2) aids in planning training, but it clusters together languages whose common features may cause Americans difficulties in learning, yet whose nature can differ radically in structure and thought patterns from language to language. Thus, while languages can be grouped together depending on how much time they demand (the current basis of the government's lists grouping languages for difficulty), languages can also be grouped depending on the kinds of difficulties they involve. For example, the four hardest languages for native English-speaking Americans to learn— Arabic, Chinese, Japanese, and Korean— share the difficulty of different writing systems and of non-Western European culture, but from that point on there are more divergencies than commonalities. For example, Chinese is a tone language. Chinese, Japanese, and Korean make a wide variety of sociolinguistic distinctions. Moreover, some languages are agglutinative, like Eskimo or Turkish, in which *evlerden* = *ev* + *ler* (plural marker) _ *den* "from" (Preposition) = "from the houses," others isolating (analytic) like Chinese or Vietnamese, still others inflecting (synthetic) like Slavic languages.[6] Numerous taxonomies could be devised to categorize these differences. The point is that identifying the level of difficulty is not sufficient. Do some learners have an affinity for, say, languages that use noun compounds versus those that use prepositional phrases to describe the same object, e.g., *Schreibmachine* (German), *skrivmaskin* (Swedish), *ritvél* (Icelandic) versus *machine à écrire* (French), and *máquina de escribir* (Spanish), all meaning "typewriter"? What about patterns of thought? With French striving for clarity and concision, while German essays the overarching sentence that perfectly qualifies the topic before it is finally mentioned, we have two truly differing ways of writing about the world.

Q5: In What Skill Modalities Will the Examinee Excel?

In other words, shouldn't aptitude tests be sensitive to modalities? Not everyone writes his or her native language as well as he or she speaks it. Nor does everyone in the world who speaks a language, read it. These facts suggest that while there are four skill modalities—

Table 2
DLIFLC Categories of Language Difficulty (Selected Languages)

Category 1	Category 2	Category 3	Category 4
Afrikaans	German	Albanian	Arabic
Danish	Hindi	Amharic	Chinese-
French	Idonesian	Armenian	Mandarin
Haitian-	Malay	Azerbaijani/	Cantonese
Creole	Rumanian/	Azeri	Japanese
Italian	Moldavian	Bashkir	Korean
Norwegian	Urdu	Belarussian	
Portugese		Bengali	
Spanish		Bulgarian	
American		Cambodian	
Caribbean		Czech	
Castilian		Estonian	
Creole		Finnish	
Swahili		Georgian	
Swedish		Greek	
		Hebrew	
		Hungarian	
		Kazakh	
		Laotian	
		Latvian	
		Lithuanian	
		Macedonian	
		Persian-Farsi	
		Polish	
		Russian	
		Serbian-Croatian	
		Slovenian	
		Somalian	
		Tadzhik	
		Tagalog	
		Tatar	
		Thai	
		Turkish	
		Turkmen/Turkoman	
		Ukrainian	
		Uzbek	
		Vietnamese	

Listening, Reading, Speaking and Writing—one person may not control all equally well. Which skill modalities does the government need? How well will a given student learn them? Should a reticent student be trained to speak? A voluble one to write? And so on. Can an aptitude test provide any clues about which skill is most likely the examinee's best or worst?

Q6: How Well Will the Examinee Attain the Course Goals?

Generally, government language courses have a goal and the jobs of graduates have a designated level to show how well the language must be controlled to do the job. For example, at DLIFLC exit scores of 2 in all skills but writing are the minimum target for its basic language programs; at FSI the desired goals are 3 in Speaking and 3 in Reading; while at the Central Intelligence Agency (CIA) a 3 in Listening, Speaking, and Reading are the desired outcomes, with Speaking paramount. Will the student achieve these levels? Can the test provide any predictive clues? Can one predict when a person will plateau? How often will they plateau? Can one predict the frequency of such plateaus?

A further complication in answering these questions stems from a language's relative difficulty for American-born English speakers. The difficulty list masks another problem: namely, language difficulty may change depending on the level we wish the examinee to achieve.

Spanish is often viewed as an easy language. In the government lists it appears as a Category 1 language; that is, among the easier to learn. Yet to put imperatives in Spanish in the negative requires subjunctive forms, and many of the tasks at Level 3 require both the present and past subjunctive. An even more striking example is Indonesian, a Category 2 language. This language is regarded among the easier for Americans to learn. However, few Americans cross the 2+/3 border. Why? A closer look suggests that low on the ILR scale Indonesian is relatively easier to learn, but harder to acquire in the mid and upper ranges. This appears to be due to two interconnected features, morphology and a syntactic feature known as the object-oriented sentence. Neither hurdle is insurmountable, but the main obstacle appears to be that Americans have difficulty knowing those contexts where a native Indonesian speaker would employ the object-oriented sentence.

In other words, besides the morphology and the object-oriented syntactic pattern, the non-native must acquire a new and rather pervasive way of thinking. As a consequence, Indonesian might be regarded as a Category 1 language if one wishes to achieve any level between 0 and 2+, but proves to be a Category 2 language if one wants

to go beyond 2+ into 3, 3+, 4, 4+, or 5! Can an aptitude test predict success for languages of shifting degrees of difficulty such as Indonesian?

Q7: How Far Can a Person Ultimately Go in Learning a Language?

While there may be an immediate exit goal for the course, what indicators can an aptitude test furnish about what the ultimate end point in the person's life-long learning of the target language? In other words, for what level, in what skill, and in what language does the examinee have aptitude?

Most of these questions target areas of ability beyond the information current aptitude tests provide, and, consequently, they extend beyond what has previously been construed as "language aptitude." While these questions have not generally been posed outside government, they logically follow from an expanded construct of language aptitude. If you can learn a language (Q1), what language best matches your personality (Q2), how difficult can it be compared to American English (Q3), and connected with what type of language: isolating, agglutinative, inflecting (Q4), in what skill modalities (Q5), to what level in the course (Q6), and ultimately to what level later in life (Q7)?

An analytic approach might regard these as seven separate questions targeting separate abilities. But a holistic view regards them as ever more precise formulations of the predictive power of a single overall ability — "to learn and to use another language." How can one design tests that will take us further along the path(s) of answering these questions than current language aptitude tests do?

Visualizing Current and Future Aptitude Tests

Visual representations of the focus and the search may be helpful in clarifying what current and future tests may require that is different or that differ from their configuration of tasks. To this end, the following figures posit differing possible solutions to aptitude test design. Although we discuss such designs in a highly abstract manner, we make reference to extant tests.

First, we simplify matters by looking at a single task and its effect on determining aptitude. This is simpler than the reality, but permits us to sketch several approaches in a concise manner before introducing reality's complications.

Let us suppose that ascertaining a person's ability to learn and use a foreign language were determinable by a single task and that task would target the speaking skill modality. We choose speaking because it is learned in the native language before reading and writing, and moreover, because it is a production skill (see Figure 2). How would we proceed? Could a speaking test in one's language determine ability to learn and use a second language? If so, what skill modality?

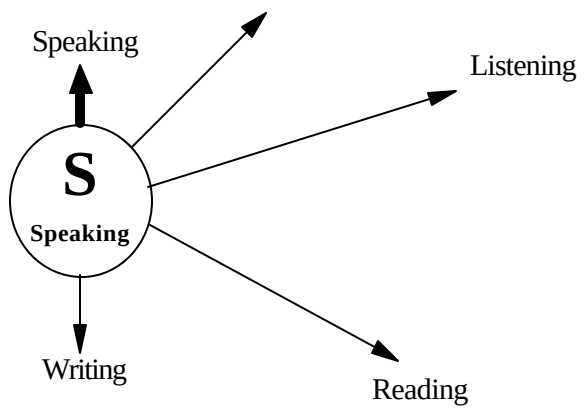

Figure 2. "One size fits all": One skill modality is the predictor

While we are operating here with single tasks, the points made about them could be readily expanded to include any range of tasks for the skill modality the single task represents. Hence, instead of a single speaking task, we could conceive of a battery of two or more tasks, and extrapolate results from the battery rather than from the single task.

Assume for a moment, however, that we found the one speaking task telling us that a person could speak well. Note it would indicate some ability to learn a language. But would it predict the ability to read, to write? Presumably it would include the ability to listen. But how well? (See Lowe, 1985, for the "offset" problem.) Since people learn to read and write after they learn how to speak and to listen, and some never learn to read, and still others never learn to write, using a single speaking task won't work! Obviously a single task could not predict in those cases where the other skill modalities, reading and writing, were never learned! Moreover, even if the chosen task could predict an

ability to learn and to use those skill modalities, the "offset" problem though not fully worked out for all the skill modalities would resurface. Think of the "offset" between one's ability to speak, listen, and read, versus one's ability to write. The last, in the U.S. at least, is usually lower than that of the other skills. So a single speaking task might indicate an ability to learn to speak and indicate some "general language ability," but it might well not predict ability to learn the other skills.

What about a single "general language factor," Oller's unitary competence hypothesis? Oller based this hypothesis on the "g" factor in intelligence. Oller has been inclined to equate the two, "g" and a "general language factor (glf)." Subsequently, he withdrew the hypothesis, having learned that his proof failed for various statistical reasons (Oller, 1980, cited in Vollmer, 1983). Still, I separate them in this article.

The unitary competence factor, however, has continued its existence anyway (Hughes & Porter, 1983). Why won't it die? Well, it is possible to adduce evidence for Oller's position up to a point, indicating that there is doubtless a strong glf, but that it is simply not as strong as "g" in intelligence. Contemplate the following situation: You give a well-designed listening comprehension test at ILR Level 3. One particular examinee does exceedingly well on the test. Your reaction might well be, not only that he understood the language at the level in question, but that the test in fact indicates that he really knows the language rather well. Yet, at the same time, you could in no way state that he also *writes* the language equally well; that is, tested listening ability cannot predict ability in another skill. You must test separately for that. I have not found it possible to predict language learning ability in a particular skill modality unless the aptitude test contained *tasks targeted against that modality*. Hence, one cannot predict a student's level of reading ability from the level of speaking ability. The import for aptitude test design is that we should have a separate listening, reading, speaking, or writing component or even separate tests, if we desire to make statements about a given skill modality.

There is in our abstract approach a possible single task solution involving "g" or "glf" (see Figure 3). This might well be a cognitive task that would determine general language aptitude and, ideally, ability in the other skill modalities as well. If one equates "glf" with "g," then the task could be drawn from intelligence testing. If one keeps "glf" separate from "g," then the task to be chosen would be taken from those thought to represent "glf" that do not connect with general intelligence tasks, and would therefore predict the examinee's degree of "glf."

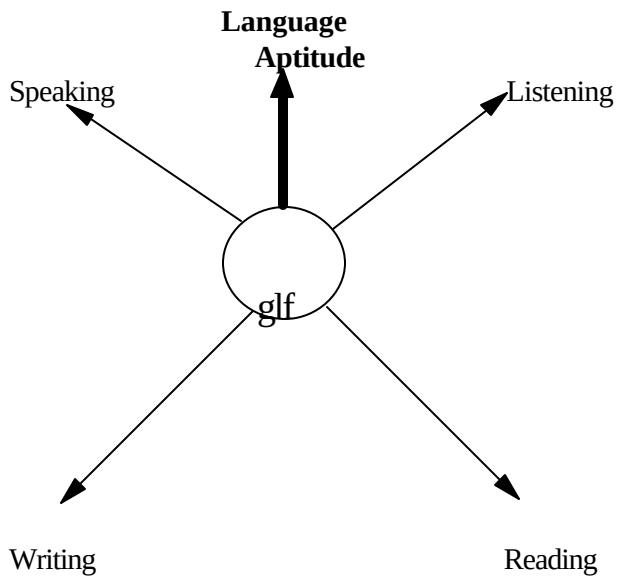

Figure 3. "One size fits all": glf is the predictor

In the total test, tasks representing "glf" need to be supplemented by tasks representing any single skill modality to ascertain the extent of possible success in that skill modality. Every skill modality has a set of related subskills that must be tested in order to ascertain its full ability. Which subskill(s) must be tested is an open question. With these thoughts we leave the "one size fits all" philosophy of ascertaining language aptitude (either that of a single skill modality or that of "glf").

The opposite pole is the “Mixed Menu” philosophy: “One from column A, one from Column B.” Staying again with a single task per skill modality, we could have one task for predicting ability, one task for listening, one for reading, one for speaking, and one for writing. (See Figure 4). The question to be answered is: Does the whole equal the sum of the parts? Does predicting success in each of the four skills provide some entrée into overall language aptitude? A variation on this solution would introduce into the previous set of tasks one task for predicting the "glf." Again, does the whole (a fuller whole this time) equal the sum of more of its parts? (See Figures 5 & 6.)

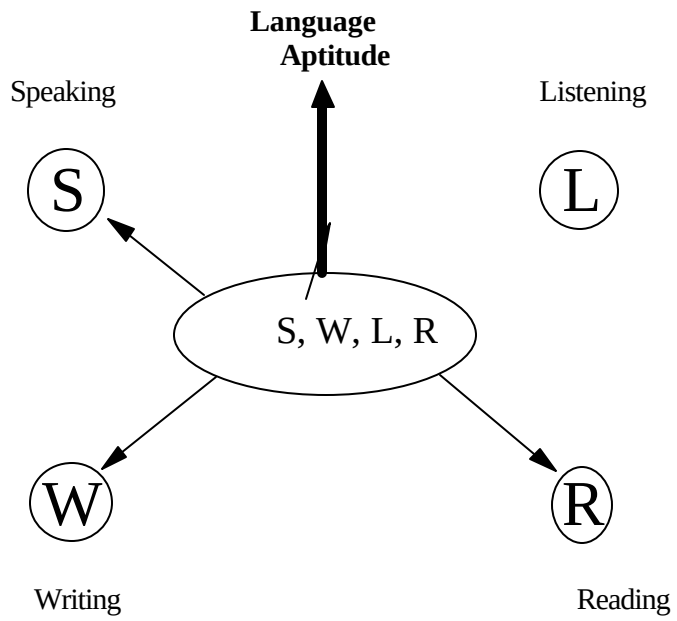

Figure 4. "Mixed menu": One from column A, one from column B: Parts predict whole

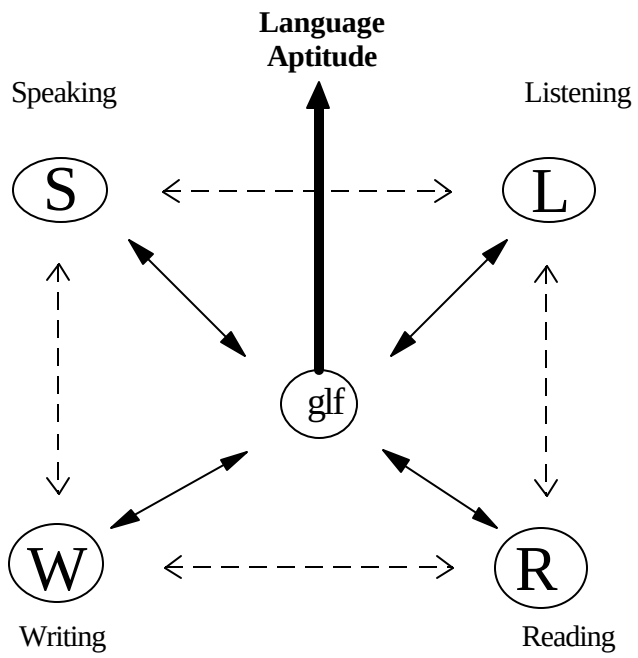

Figure 5. "Mixed menu": One from column A, one from column B

Considering the extant aptitude tests, we can say that their designs more fully resemble the "Mixed Menu" solution, often with multiple tasks. Normally, we choose these tasks by such statistical methods as factor analysis. We try a number of tasks possibly indicating "language aptitude." Then, through factor analysis we identify those that load most on the trait. Subsequently, the ones with the highest loadings are selected for the aptitude test.

In reviewing the current tests, one is led to the questions: What are these tests focusing on? Which of the questions posed earlier does a given aptitude test attempt to answer? With what success?

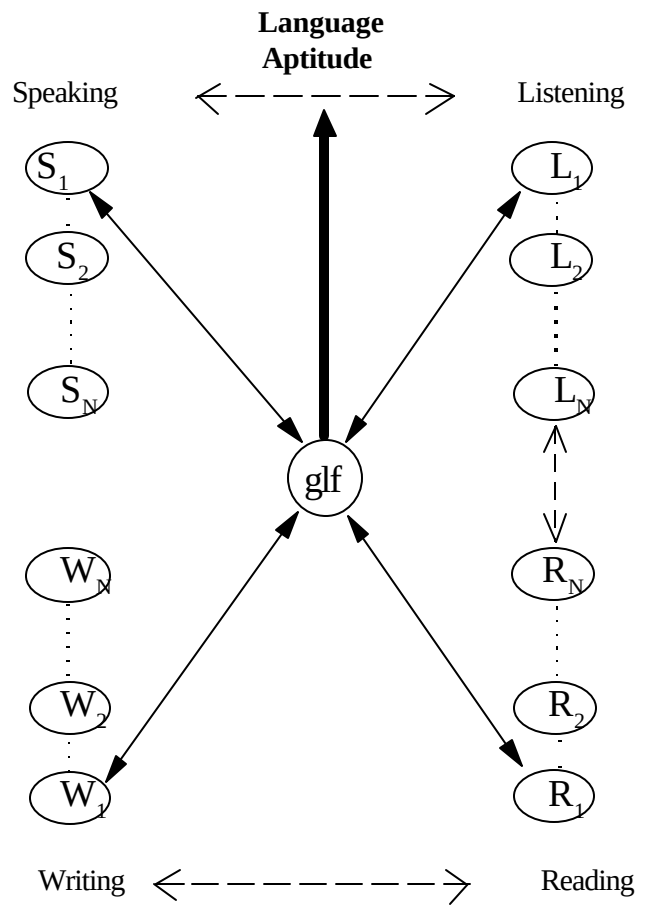

Figure 6. "Mixed menu": One from column A, one from column B: Parts predict the whole!

Current tests appear to answer our first question with some degree of accuracy: i.e., a correlation of .50 between score on the MLAT and exit proficiencies. Table 3, however, suggests that correlations between DLAB scores and exit proficiencies vary for individual skill modalities and test design.This variance again raises the question, what is the focus of the test?

Table 3
Correlations Between DLAB scores and Outcome Variables by Language by Year: Russian

YEAR	STATISTIC	DLPT-L	DLPT-R	DLPT-S
1986	r	0.29077	0.40175	0.12764
	p	0.0001	0.0001	0.0006
	n	729	729	729
1987	r	0.34892	0.41579	0.22475
	p	0.0001	0.0001	0.0001
	n	567	566	567
1988	r	0.29463	0.37770	0.17589
	p	0.0001	0.0001	0.0001
	n	636	636	636
1989	r	0.29064	0.31847	0.22066
	p	0.0001	0.0001	0.0001
	n	834	834	834
1990	r	0.31858	0.32597	0.20292
	p	0.0001	0.0001	0.0001
	n	833	833	832
1991	r	0.31147	0.36340	0.18500
	p	0.0001	0.0001	0.0001
	n	796	796	796
1992	r	0.20546	0.22894	0.17374
	p	0.0001	0.0001	0.0001
	n	824	824	824
1993	r	0.22952	0.28561	0.22778
	p	0.0001	0.0001	0.0001
	n	678	678	678
1994	r	0.28512	0.46500	0.22988
	p	0.0001	0.0001	0.0002
	n	255	255	255

Since the current aptitude tests are paper-and-pencil only tests, they automatically involve reading. Some have audiotaped sections, but do such sections really test listening or is their primary goal to test memory span? All these tests are machine-scorable so they cannot test writing. And none contain any spoken language tasks, so they cannot test the ability to learn speaking! What are we testing? What is the concept of aptitude that underlies the test? What can we say as a result

of the test? Perhaps, in retrospect, we should wonder that we can say as much as we do at least as regards general aptitude.

Conclusion

Focusing on a future aptitude test design, should we:

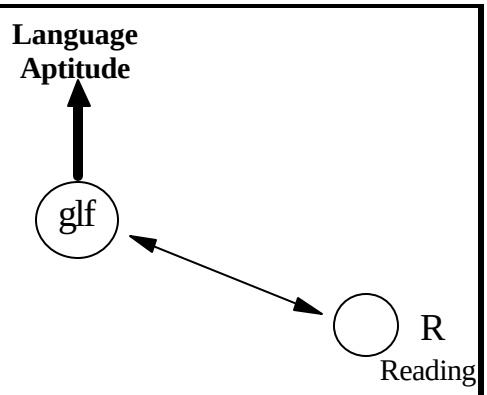

Figure 7. "Mixed menu": One from column A, one from column B: General language aptitude

- attempt to test for "glf " to determine those who would and
- those who would not benefit from language training?
- target aptitude tests against against a single skill modality, for example, reading (see Figure 7)?
- give targeted aptitude tests against languages of a specific structure such as Child's VORD with its Turkic-based artificial language?
- attempt to predict how far the examinee will go in learning the language?

These questions could be multiplied, but their implication is clear. In the future, what will be our definition of the "language aptitude" construct? And more importantly, what will be the focus of our language aptitude tests (Q0)? Which of these questions can a language aptitude test conceivably answer satisfactorily? Which only partially, and which not at all?

Notes

[1] The views expressed herein are those of the author and in no way represent those of the Department of Defense.

[2] I am indebted to both Ray T. Clifford and James R. Child for critiquing earlier drafts of this article. The latter was particularly helpful in the revision of the earlier oral version, from September 1994, into this written one. I regret only that I have been unable to reflect all of their comments. Naturally, any remaining errors are the responsibility of the author. Finally, I also wish to thank John Lett for providing the correlation statistics from ongoing research at DLIFLC.

[3] VORD is not an acronym. Child has been so plagued by questions as to its meaning that he ultimately resorts to saying, "No, it's not an acronym! But if you must confer meaning on it, then I'll say that it's the word for 'word' in VORD."

[4] Ray T. Clifford (through personal communication) points out that "... .50 is actually very high compared to other attempts at predicting human behavior over an extended period of time." My point is rather: Can we say more about less, which I hope will become clear through the designs suggested later in the article?

[5] After Q1 I am not exactly sure as to the order of the following questions.

[6] Bernard Comrie. (1981). *Language Universals and Linguistic Typology.* Chicago, IL: University of Chicago Press

References

Artificial Language Aptitude Test (ALAT). (1951). Washington, D.C.: Department of the Army.

Carroll, J. B., & Sapon, S. M. (n.d.) *Elementary Modern Language Aptitude Test* (MLAT).

Carroll, J. B., & Sapon. S. M. (1959). *The Modern Language Aptitude TEST* (MLAT). New York: The Psychological Corporation.

Child, J.R. (1973). VORD. Washington, DC: Department of Defense.

Comrie, B. (1981). *Language universals and linguistic typology*. Chicago, IL: University of Chicago Press.

Defense Language Aptitude Battery (DLAB). (1977). Monterey, CA:
Defense Language Institute Foreign Language Center.

Lowe, Jr., P. (1985). The ILR Scale as a synthesizing research principle. In C. James (Ed.), *Foreign language proficiency in the*

classroom and beyond. Lincolnwood, IL: National Textbook. P. 35ff.

Parry, T. S., & Child, J. (1990.) Preliminary investigation of the relationship between VORD, MLAT, and language proficiency, In T. S. Parry & C. W. Stansfield (Eds.). *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall Regents. Pp. 30-66.

Petersen, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement, 6*, 369-380.

Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. New York: Harcourt Brace Jovanovich.

Vollmer, H. J. The structure of foreign language competence," In A. Hughes & D. Porter (Eds.), *Current developments in language testing*. London: Academic Press.

Author

PARDEE LOWE, JR. Senior Language Analyst, National Cryptologic School, Department of Defense. Specializations: aptitude, oral proficiency testing, passage grading, performance, and proficiency testing.

Applied Language Learning
1998, Vol. 9, Nos. 1 & 2, pp. 31-70

The Modern Language Aptitude Test for Predicting Learning Success and Advising Students

Madeline Ehrman
Foreign Service Institute

The Modern Language Aptitude Test (MLAT) was part of a project examining biographical, motivational, attitudinal, personality, and cognitive aptitude variables among 1,000 adult students preparing for overseas assignments at the Foreign Service Institute (FSI) with various smaller numbers for sub-samples completing different instruments. Data were analyzed by correlation, Analysis of Variance (ANOVA), chi-square, and multiple regression as appropriate to the data and the research questions. The MLAT proved the best of the available predictors of language learning success. As part of an effort to expand the concept of language learning aptitude beyond strictly cognitive factors, this study relates the MLAT not only to end-of-training proficiency outcomes, but also to personality dispositions, using both overall correlational data and information on extremely strong and weak learners. Qualitative findings from use of the MLAT part scores in student counseling activities are also described, suggesting utility for this instrument beyond prediction of learning success.

This article describes findings of research in progress at the Foreign Service Institute (FSI), a U.S. government language training institution. For years, incoming students have taken the MLAT; indeed, a sample from FSI was among the groups on which the MLAT was originally normed (Carroll & Sapon, 1959). It is still used as part of the institute's procedures for assignment to foreign language training. (Language aptitude testing is also done at other agencies.)

© 1998, Madeline Ehrman

Over recent years, the MLAT has become the subject of some controversy at FSI. Some program managers continue to see a good relationship between performance on the MLAT and in language training. Others protest that the relation, such as it is, is not very strong and furthermore the MLAT may not represent the true ability of those who lack formal education (Rockmaker, personal communication, 1993). Anti-MLAT opinion has also suggested that the MLAT was designed for the audio-lingual methodology that was in vogue in the late 1950s and 1960s and that the test is no longer valid for the much more "communicative" teaching that is now done at FSI (Bruhn, personal communication, 1992). Much of the distrust of the MLAT is connected with the increased suspicion of psychological testing during the last quarter century (Anastasi, 1988). The project on which this paper reports was initiated in order to take such concerns about the MLAT out of the realm of allegation and find out just how useful it still is.

The present article reports on two efforts to addressthese concerns. One is a quantitative investigation of a large sample of FSI students between 1992 and 1994. That study looks at the MLAT primarily as a predictor of language learning success in the FSI setting of intensive, full-time language learning for communicative use. The other portion of the article describes a less rigorous attempt to make use of patterns of high and low MLAT part scores with individual students. The initial outcomes of this attempt, still highly exploratory, suggest that the MLAT may have value for pinpointing areas of learning success and difficulty for a wide range of students, including some relatively able but context-dependent ones not well served by relatively grammar-oriented instruction.

Review of Literature

The MLAT was perhaps the culmination of a long tradition of psychometric test development and efforts to predict language learning achievement. It achieved a fairly respectable level of success in the audio-lingual and grammar-translation classrooms of the 1950s and 1960s (Spolsky, 1995). Other important language aptitude tests developed out of the same tradition include the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966), the Defense Language Aptitude Battery (DLAB) (Petersen & Al-Haik, 1976), and VORD (Parry & Child, 1990). The Pimsleur is different from the MLAT in particular because it includes a portion directly addressing the ability to infer language structure from an artificial language stimulus. The DLAB consists primarily

of such induction-testing items, in a modified English. VORD was designed to test the ability to cope with the grammar of languages in the Altaic family and consists of items that test such grammatical prowess (Parry & Child, 1990). All four, including the MLAT, were found to have similar predictive validity (Parry & Child, 1990). This article will not address these other instruments, but will focus on the MLAT, which is the instrument that is still in use at the Department of State.[1]

The outcome of a major research project at Harvard University, the MLAT is based on a factor analysis of a large number of individual characteristics thought to contribute to language learning. Carroll (1962) describes the project in extensive detail; the MLAT Manual (Carroll & Sapon, 1959) provides information on the validation studies. The individual characteristics were grouped into four main categories: phonetic coding ability (distinguishing sounds and reflecting them graphically), grammatical sensitivity (recognizing and using syntactic relationships), memory (rote and contextualized), and inductive language learning. All but the last of these four are directly addressed in the five parts of the MLAT (see Appendix A).

Other components listed by scholars of language aptitude include motivation and knowledge of vocabulary in the native language (Pimsleur, 1968), the ability to hear under conditions of interference (Carroll, 1990), the ability to "handle decontextualized language" (Skehan, 1991), and the ability to shift mental set and cope with the unfamiliar (Ehrman, 1994b, 1995b, 1996; Ehrman & Oxford, 1995).

A desire for better prediction of language learning and the ability to exploit aptitude testing further has led to recent research efforts. At least two major projects in recent years have examined the role of individual differences in addition to strictly cognitive aptitude in language learning. They are the Defense Language Institute Foreign Language Center's (DLIFLC) Skill Change Project (Lett & O'Mara, 1990) and FSI's Language Learning Profiles Project (Ehrman, 1993, 1994, 1995b, 1996; Ehrman & Oxford, 1995; Oxford & Ehrman, 1995). Both investigated such variables as biographic factors, personality, motivation, anxiety, and learning strategies, as well general intelligence (DLIFLC only). A similar project was begun at the Central Intelligence Agency language school, though without personality variables, and DLIFLC engaged in a large-scale effort to review the DLAB (Thain, 1992; Lett & Thain, 1994). This article is part of the project at FSI.[2]

Across a number of studies, predictive validity correlations for the MLAT have generally ranged between .42 and .62 for most

languages, with outliers of .27 for certain non-Indo-European languages at the DLIFLC and as high as .73 with language instructor ratings of student performanced at FSI (Carroll & Sapon, 1959). More recent studies of the MLAT produce quite mixed results. Brecht, Davidson, and Ginsburg (1993) did not find the MLAT predictive of overall oral proficiency in intensive language training in Russian. However, for the same programs they found Part 3 (Spelling Clues) to be "highly significant" in predicting listening comprehension and the Total Score to be significantly predictive of reading proficiency. They speculate that the complex nature of the communicative task causes the lack of predictive value for oral proficiency. This suggestion is quite consistent with the questions raised at FSI (see above) and the point of view that standard aptitude measures do not "take into account" such developments as focus on communicative competence, pragmatics and discourse, new thinking by cognitive psychologists (Parry & Stansfield, 1990).

Another finding is that of Spolsky (1995), who reports that MLAT Part 1 correlated significantly with success on the part of Israeli learners of French as a foreign language, but the MLAT did not predict achievement in Hebrew at the same school. He suggests that this variance may be related to differences in such factors as motivation, which is so powerful that it may override aptitude. (I suggest that it may also be the case that the students were learning Hebrew as a second language, not a foreign language, so not all of their learning was classroom-based, which is the task for which existing language aptitude tests were designed.)

Most of the research cited addresses the use of the MLAT (and other aptitude measures) as predictors of learning success, and indeed this is an important consideration for assignment to intensive and long-term language training at taxpayer expense. However, a measure like the MLAT also has potential utility for *placement* in a program (Wesche, 1981) and *diagnosis* of learning difficulties, for counseling students, and for tailoring programs to their needs (e.g., Demuth & Smith, 1987; Sparks, Ganschow, & Patton, 1995). These applications have received far less attention in the literature. They are also among the areas of interest for the FSI investigation, and it is in these that the MLAT has been successfully used (Lefrancois & Sibiga, 1986; Wesche, 1981).

Methods

Sample

In this study, there are 343 students altogether with at least a single MLAT score; of these, part scores for the five subscales are available for 296. Males constitute 59% and females 41% of the sample. The average age of students at the time of participation was 39, with a standard deviation of 9 years. The median education level was between bachelors and masters degrees. Of those that report previous language study, the average number of languages studied was 1.8.

In the presentation of correlations with other instruments, numbers are smaller because not every person in the data set with an MLAT score completed all the other instruments. For example, of the 343 students with at least one MLAT score, only 93 had scores on the Myers-Briggs Type Indicater (MBTI) Form G. On Tables 4 and 5 in the results section of this article, which are excerpted from another sub-study in the FSI Language Learning Profiles Project , the numbers are different from those in the present study, though they represent overlapping subsamples from the same population of students.

FSI trains and tests students not only from its parent agency, the Department of State, but also from many other agencies. Students from the Department of State comprise 70% of FSI language students. Other agencies sending the most students are the United States Information Agency, the Department of Defense, the Department of Commerce, and the Agency for International Development.

Students in this study are beginners in long-term (i.e., 16 weeks or longer) intensive language training. The languages they are studying are classified into four categories based on agency experience with the length of time needed by English speakers to reach "professional" proficiency (S-3, R-3—see 'Instrumentation' for a brief description of the ILR rating scale):

1. Western European;
2. Non-Western European but relatively quick for English speakers to learn (Swahili, Indonesian, and some North European languages);
3. Other non-Western European (e.g., Russian, Thai), but excluding the Category 4 languages;
4. "Superhard" languages (Arabic, Chinese, Japanese, and Korean).[3]

Usual training lengths vary by language category. Most FSI students are expected to reach "professional" proficiency (S-3 R-3) in 24 weeks in a Category 1 language, in 32 weeks in a Category 2 language, in 44 weeks in a Category 3 language, and in 88 weeks (2 academic years) in a Category 4 language.[4] These expectations are normally reflected in the lengths of student assignments to training and are also accounted for in the statistics reported in this article.

Instrumentation

The MLAT (Carroll & Sapon, 1959) is the classic language aptitude test, with 146 items. The MLAT Manual describes its five parts: 1. Number Learning (memory, auditory alertness); 2. Phonetic Script (association of sounds and symbols); 3. Spelling Clues (English vocabulary, association of sounds and symbols); 4. Words in Sentences (grammatical structure in English); and 5. Paired Associates (memorizing words), together with a total score. The MLAT was correlated .67 with the Primary Mental Abilities Test (Wesche, Edwards, & Wells, 1982), suggesting a strong general intelligence factor operating in the MLAT. Split-half reliabilities for the MLAT are .92 - .97, depending on the grade or age. For college students, validity coefficients (correlations with course grades) provided in the MLAT Manual (Carroll & Sapon, 1959) are .18 - .69 for the long form of the MLAT and .21 - .68 for the short form. For adult students in intensive language programs, validity coefficients (correlations with teacher ratings) in the Manual (Carroll & Sapon, 1959) are .27 - .73 for the long form and .26 - .69 for the short form. This study used the long form.

The subscales of the MLAT are described briefly in Appendix A. The Index Score used at FSI originated in the 1960s as a T-score based on the Total score, with three standard deviations of 10 on either side of a mean of 50.[5] It has since become frozen as a translation of the Total, much like Scholastic Aptitude Test scores until recently, because of the agency personnel system's dependence on over 30 years of Index records. For users of the MLAT who are more familiar with the raw Total score, a table of equivalences is provided in Appendix B.

Note that an Index of 50 is the mean established when the MLAT was originally normed and includes a variety of subjects from high schools and colleges. Whether it in fact is still representative of the population outside FSI is uncertain. What is certain, however, is that a mean Index of 50 is no longer valid for FSI students. There has been a

gradual upward tendency in the MLAT Index mean at FSI over the intervening 30 years. Wilds (1965) reported a mean Index of 54 (N=957, no SD); an agency-internal document reports a 1984 mean Index of 59, SD 10, N-312 (Adams, 1984); and the mean Index for all the students in the current sample who had MLAT scores is 63, SD 10, N = 343. [6]

End-of-Training Proficiency Tests

These tests provide the main criterion measure in this study. At the end of training, FSI students are given proficiency assessments resulting in ratings ranging from 0 to 5 for speaking (the S-score, which includes interactive listening comprehension) and for reading (the R-score). The full oral interview, including speaking, interactive listening, and an interactive reading test using authentic material, takes two hours. R-3, for example, indicates reading proficiency level 3 ("professional" proficiency); S-2 represents speaking proficiency level 2 (working proficiency). Other levels are 0 (no proficiency), 1 (survival level), 4 (full professional proficiency, with few if any limitations on the person's ability to function in the language and culture), and 5 (equivalent to an educated native speaker). "Plus" scores (e.g., indicating proficiency between S-2 and S-3) were coded as 0.5; thus, for example, a score of S-2+ was coded 2.5.

The ratings are equivalent to the guidelines of the ILR/ACTFL (Interagency Language Roundtable/American Council on the Teaching of Foreign Languages) that originated at FSI and have been developed over the years by government agencies. These guidelines are detailed by Omaggio, 1986. Most students enter FSI with goals of end-of-training proficiency ratings at S-3 R-3 for full-time training, comparable to ILR/ACTFL Superior Proficiency.

Reliability studies have shown that government agencies have high inter-rater reliability for proficiency ratings within a given agency, but that the standards, or their interpretations, are not always the same at every agency. Thus, raters at different government agencies do not have as high an inter-rater reliability as raters at the same agency. Proficiency ratings are therefore considered reliable indicators of the level of language performance of an individual student within an agency (Clark, 1986).

Learning Style, Strategy, and Personality Instruments

The Learning Style Profile is a pure learning style instrument: that is, it is neither a personality questionnaire nor an aptitude test. The Myers-Briggs Type Indicator and its Type Differentiation Indicator scoring system are both a personality instrument and a way to assess learning style, as is the Hartmann Boundary Questionnaire. The student learning activities questionnaires tap learning strategies.

The Hartmann Boundary Questionnaire (HBQ, Hartmann, 1991) was developed for research with sleep disorders and nightmares, using a psychoanalytic theoretical base. It is intended to examine the degree to which individuals separate aspects of their mental, interpersonal, and external experience through "thick" or "thin" psychological boundaries. Its 146 items address the following dimensions: sleep/dreams/wakefulness, unusual experiences, boundaries among thoughts/feelings/moods, impressions of childhood/adolescence/adulthood, interpersonal distance/openness/closeness, physical and emotional sensitivity, preference for neatness, preference for clear lines in pictures or clothing, opinions about children/adolescents/adults, opinions about lines of authority, opinions about boundaries among groups/peoples/nations, opinions about abstract concepts, plus a total score for all twelve of the above scales. For example, thin boundaries are represented by commonly drifting in and out of sleep states while waking up, memory for experiences at a variety of ages, tolerance for lack of order in the workplace, or preference for little organizational hierarchy. Thick boundaries are suggested by the opposite approach, e.g., preference for thick, heavy clothing, interpersonal distance, or beliefs that children should be seen and not heard. Hartmann found women and younger people score consistently "thinner" than men and older people. Cronbach alpha reliability for the HBQ is .93, and theta reliabilities for subscales are .57 - .92 (Hartmann, 1991).

The National Association of Secondary Schools Principals' Learning Style Profile (LSP), (Keefe & Monk, with Letteri, Languis, & Dunn, 1989) is a 125-item composite measure composed of many different approaches to measuring learning style. The main subscales are cognitive skills (analytic, spatial, categorization, sequential processing, detail memory, discrimination), perceptual response (i.e., sensory preferences: visual, auditory, emotive/kinesthetic), orientations (persistence, verbal risk-taking, manipulative), study time preferences (early morning, late morning, afternoon, evening), and environmental context for learning (verbal vs. spatial, posture, light, temperature, mobility, and

grouping). Cronbach's alpha for the subscales ranged from .47 to .76, with an average of .61. Test-retest reliabilities were .36 to .82 after 10 days and somewhat lower after 30 days. Concurrent validity of the LSP's analytic subscale with the Group Embedded Figures Test was .39. Concurrent validity of the perceptual response subscales of the LSP with the Edmonds Learning Style Identification Exercise was .51 - .64. Many of the environmental context subscales of the LSP correlated with Dunn and Dunn's Learning Style Inventory, .23 - .71. All concurrent validity scores are reported in the manual with a significance value < .002.

The Myers-Briggs Type Indicator (MBTI, Myers & McCaulley, 1985), Form G, is a 126-item, forced-choice, normative, self-report questionnaire designed to reveal basic personality preferences on four scales: extraversion-introversion (whether the person obtains energy externally or internally), sensing-intuition (whether the person is concrete/sequential or abstract/random); thinking-feeling (whether the person makes decisions based on objective logic or subjective values); and judging-perceiving (whether the person needs rapid closure or prefers a flexible life). Internal consistency split-half reliabilities average .87, and test-retest reliabilities are .70 - .85 (Myers & McCaulley, 1985). Concurrent validity is documented with personality, vocational preference, educational style, and management style (.40 - .77). Construct validity is supported by many studies of occupational preferences and creativity.

The Type Differentiation Indicator (TDI) (Saunders, 1989) is a scoring system for a longer and more intricate 290-item form (MBTI, Form J) that provides data on the following subscales for each of the four MBTI dimensions: extraversion-introversion (gregarious-intimate, enthusiastic-quiet, initiator-receptor, expressive-contained, auditory-visual); sensing-intuition (concrete-abstract, realistic-imaginative, pragmatic-intellectual, experiential-theoretical, traditional-original); thinking-feeling (critical-accepting, tough-tender, questioning-accommodating, reasonable-compassionate, logical-affective); and judging-perceiving (stress avoider-polyactive, systematic-casual, scheduled-spontaneous, planful-open-ended, methodical-emergent). The TDI includes seven additional scales indicating a sense of overall comfort and confidence versus discomfort and anxiety (guarded-optimistic, defiant-compliant, carefree-worried, decisive-ambivalent, intrepid-inhibited, leader-follower, proactive-distractible), plus a composite of these called "strain." Each of these comfort-discomfort subscales also loads on one of the four type dimensions, e.g., proactive-distractible is also a judging-

perceiving subscale. There are also scales for type-scale consistency and comfort-scale consistency. Reliability of 23 of the 27 TDI subscales is greater than .50, an acceptable result given the brevity of the subscales (Saunders, 1989).

Student Learning Activities Questionnaires

At the beginning of training, students were asked to complete the Motivation and Strategies Questionnaire (Ehrman & Christensen, 1994), which has a variety of questions about self-efficacy as a learner, motivation, anxiety, a detailed list of classroom activities, and a set of individual activities. At the end of the students' training (minimum of 16 weeks, maximum of 44 weeks, depending on the difficulty category of the language), each participant in the study was asked to complete two questionnaires: "CLASSACT" (Ehrman & Jackson, 1992) on relative usefulness of a fairly detailed list of classroom activities (Likert scaled 1-3) and "SELFACT" (Hart-Gonzalez & Ehrman, 1992) on relative usefulness (1-3) of their own study activities and estimated time per week devoted to each. These questionnaires are used here for the first time. Because completion at the end of training was voluntary and students were very busy with preparations for departure, the return rate was low (about 10%), and numbers for several of the items are not adequate for analysis. (This and other studies using these questionnaires are part of their validation. When there are sufficient cases, they will be subjected to reliability analysis and factor analysis.)

Data Collection and Analysis

Data collection took place over a two-year period, between 1992 and 1994. Students who started at the beginning of each semester were asked to participate but could decline the invitation; under 5 % of the students who were approached chose not to participate. During the 1992-1993 academic year, all French and Spanish students (who start 10 times a year) were also invited to join the study, with the same declination rate.

All questionnaires except the MLAT were administered within the first week of training. If a student already had an MLAT record, he or she could arrange for those scores to be included in the research data set; otherwise, MLAT administration took place within the first month of the beginning of training. In this sample, almost all (95%) of the MLAT scores were current, i.e., within the previous 3 years. Proficiency tests were administered at the end of training, after (in most

cases) 24 or 44 weeks.

Data analysis in this study on Statistical Product and Service Solutions (SPSS) for Windows 5.0.1 (Norusis, 1992) used correlations, one-way analysis of variance (ANOVA), and multiple regression. Correlations of the MLAT were done with end-of-training ratings for speaking and reading proficiency (the FSI proficiency test is described above, under "Instrumentation") and with individual difference variables (see above for listing and descriptions of the instruments). The data used for the correlations between end-of-training proficiency and the MLAT Index for all language categories combined were filtered to equalize expected length of training and proficiency outcomes (that is, to make results of a language like French comparable to those of a language like Chinese).

Results

Distributions

Table 1 shows that the Index Score is somewhat higher for

Table 1
MLAT Descriptive Statistics for the Index Score

Category	*N*	Mean	SD	Range	Mode	Skewness	Kurtosis
All Students	343	63	10	21-80	70	-.973	1.392
Category 1	169	59	12	21-80	61,70	-.808	.625
Categories 2-3	120	66	8	45-80	70	-.462	-.171
Category 4	54	63	10	26-78	64	-.900	.770

Minimum possible Index: 20; maximum possible Index: 80. Category 1: Western European languages; Category 2: Swahili, Indonesian, Malay; Category 3: Eastern European and non-Western languages (except Category 4 languages); Category 4: Arabic, Chinese, Japanese, Korean.

Category 2, 3, and 4 languages than for Category 1 languages in central tendency and range (see "Sample" for definitions of these categories). The part scores follow the same pattern.

The distributions, with their high central tendencies and reduced space below the ceiling for FSI students, reflect several forms of preselection. The first is that many students have self-selected for foreign affairs careers. Most of these went through their agency's selection process. This process has already probably eliminated some of the students least likely to score well on the MLAT. Second, the MLAT Index Score is used for selection of students in FSI's parent agency's personnel system, along with other evidence of likely learning, especially evidence of previous language learning success. (Such selection is authorized in the personnel regulations for the U.S. Department of State, where it is clearly stated that evidence of learning success overrides the MLAT.)

Selection is done in the State Department's personnel system especially for non-Western European languages, for which training to the "professional" proficiency level (S-3, R-3) takes 44-88 weeks. Relatively low MLAT students (Index below 55 for Category 3 or 60 for Category 4 languages) with no other evidence of success are normally sent to Western European languages, hence Category 1 is the group where we find a relatively large range of tested aptitude.

Preselection makes analysis of the MLAT's predictive value for Category 3 and 4 languages in this sample difficult. On the other hand, in view of the expense entailed by 44-week and 88-week intensive language training, assignments personnel understandably seek every indication of likely success or lack of it, without reference to the needs of the researcher.

Other results are described under two rubrics: findings related to prediction of language learning success and findings related to diagnosis and student counseling. The former are quantitative; the latter are qualitative.

Results Related to Prediction of Language Learning Success

Correlations

Correlation coefficients for MLAT Index, Total, and part scores with S- and R-ratings range in the 40s and 50s for the MLAT when a broad range of scores is available, comparable with coefficients found originally by Carroll (1990). The Index Score tends to show higher correlations with end-of-training proficiency ratings than do the part scores or the Total. Correlations for the Index Score are shown in Table 2. Category 1: Western European languages; Category 2: Swahili,

Indonesian, Malay; Category 3: Eastern European and non-Western languages (except Category 4 languages); Category 4: Arabic, Chinese, Japanese, Korean. S-rating: speaking and interactive listening; R-rating: reading.

Table 2
Correlations of MLAT Index Score with End-of-Training Proficiency Ratings

Language(s)	r	S-rating	r	R-rating
All	.44	(N = 343)	.40	(N = 341)
Category 1	.52	(N = 169)	.55	(N = 168)
Category 2-3	.34	(N = 120)	.35	(N = 120)
Category 4	.47	(N = 54)	.34	(N = 53)

Correlations are weakest for Category 2 and 3 languages and strongest for Category 1 languages, where there is the greatest range, and the distribution of MLAT scores closely resembles a normal distribution. For Categories 1-3, correlations with reading and speaking are roughly the same. In Category 4 languages, they are stronger for speaking than for reading. This difference may be due to a smaller range in reading scores (they are much lower for beginners than in other languages), or possibly because the MLAT does not address abilities needed for reading languages that use Chinese or Chinese-type characters—three out of the four Category 4 languages.

Analysis of Variance

This investigation was done only for the entire sample, because the numbers of subjects were not sufficient for Category 2 and 3, or 4 languages separately. In a study of the extremely strong and weak students in the sample, the bottom 3 to 4 % were contrasted against all others and the top 5 to 6 % against all others. Extreme students were selected on a formula that combined length of training, relative difficulty of language by category, and end-of-training scores. There were fewer students at the low end because the very weakest may be withdrawn well before their scheduled end of training. In addition, the training staff strive to find ways to maximize each student's strengths, so that as many students as possible can achieve their training goals, which in most cases are S-3, R-3. More detail on the study of extremes,

including the selection formula, is available in Ehrman (1994b).

Data for the individual difference variables were analyzed using the one-way analysis of variance procedure in SPSS for Windows 6.1. Levene's Test and t-tests for equality of means were used to determine unequal variances and the appropriate significance level, depending on whether variances were equal or unequal. The findings for the MLAT are displayed in Table 3.

Table 3
Performance Extremes: ANOVAs

Weakest, Speaking *N* selected (weakest): 4 (Parts & Total), 6 (Index)
N not selected (all others) = 292 (Parts & Total), 337 (Index).

Part	Weakest Mean	All Others Mean	Weakest SD	All Others SD	F	Sig.
1	24.5	36.5	6.5	9.1	6.8524	.0093
2	18.5	24.7	3.5	4.5	7.3634	.0070
3	11.0	28.3	8.6	9.9	12.1415	.0006
4	15.3	28.0	5.3	7.5	11.4289	.0008
5	11.5	19.3	4.7	5.3	11.4289	.0008
Total	80.8	136.7	24.6	27.5	16.3881	.0001
Index	43.2	62.7	10.8	10.5	20.5548	

Strongest, Speaking *N* selected (strongest): 14 (Parts & Total), 19 (Index)
N not selected (all others) = 281 (Parts & Total), 324 (Index).

Part	Strongest Mean	All Others Mean	Strongest SD	All Others SD	F	Sig.
1	40.5	35.0	4.9	9.7	4.4395	.0362
2	27.1	24.3	2.8	4.7	5.2765	.0225
3	32.8	27.0	7.0	14.2	4.5701	.0336
4	30.0	27.2	5.0	7.9	1.7067	.1927
5	20.8	18.8	4.2	5.5	1.6950	.1942
Total	151.2	132.5	13.8	29.6	5.7291	.0175
Index	68.2	60.9	5.9	11.2	7.8286	.0055

Weakest, Reading *N* selected (weakest): 3 (Parts & Total), 4 (Index)
N not selected (all others) = 292 (Parts & Total), 337(Index).

Part	Weakest Mean	All Others Mean	Weakest SD	All Others SD	F	Sig.
1	36.4	7.0	9.1	6.4559	.0115	
2	23.0	24.7	3.8	4.5	7.1481	.0079
3	17.7	28.2	5.5	9.9	13.4109	.0003
4	7.3	28.0	3.5	7.5	11.8901	.0006
5	11.0	19.3	5.6	5.3	7.3757	.0070
Total	72.0	136.6	21.2	27.6	16.3758	.0001
Index	40.5	62.7	12.6	10.5	17.6391	.0000

Strongest, Reading *N* selected (strongest): 78 (Parts & Total), 93 (Index)
N not selected (all others) = 217 (Parts & Total), 248 (Index).

Part	Strongest Mean	All Others Mean	Strongest SD	All Others SD	F	Sig.
1	38.9	33.8	6.3	10.5	15.0647	.0001
2	26.1	23.7	3.5	4.8	15.4653	.0001
3	31.0	26.9	8.6	10.2	14.7692	.0002
4	29.2	26.7	6.5	7.9	6.1293	.0140
5	21.3	17.9	4.1	5.6	22.5703	.0000
Total	146.5	128.0	20.9	30.1	23.7211	.0000
Index	66.3	59.6	8.0	11.3	26.1914	.0000

Data analysis done by SPSS for Windows v. 6.1, One Way Analysis of Variance Test. Degrees of freedom are available upon request.

Speaking. Of all the variables analyzed, Parts 3, 4, 5, the Total, and the Index scores best differentiated the weakest students. The MLAT variables also differentiated these weak students better than any other of the many variables in the research project.

For the strongest students' speaking scores, the Index ($F=7.83$, $.p < .0055$) was the strongest differentiator from among the MLAT and learning style variables, but it was not as good as these biographical

background variables: education level, number of previous languages, and previous highest score in speaking and especially reading (see Tables 4 and 5). The MLAT appears to differentiate the strongest speakers less clearly than the weakest speakers and the strongest and weakest readers

Table 4
Results of ANOVAs for Weakest Students Compared with All Others

	Weakest Students			Non-Weakest Students					
	N	Mean	(SD)	N	Mean	(SD)	F	df	Sig.
Category	**From the Biographic Data**								
No Prev.Lang.	15	1.0	(1.7)	674	1.7	(1.3)	7.1502	687	.008
	Modern Language Aptitude Test								
Part 1	4	24.5	(6.5)	292	36.5	(9.1)	6.8524	294	.009
Part 2	4	18.5	(3.5)	292	24.7	(4.5)	7.3634	294	.007
Part 3	4	11.0	(8.6)	292	28.3	(9.9)	12.1415	294	.0006
Part 4	4	15.3	(5.3)	292	28.0	(7.5)	11.4289	294	.0008
Part 5	4	11.5	(4.7)	292	19.3	(5.3)	8.7868	294	.003
Total Score	4	80.8	(24.6)	292	136.7	(27.5)	16.3881	294	.0001
Index Score	6	43.2	(10.8)	339	62.7	(10.5)	20.5548	343	.0000
	Hartmann Boundary Questionnaire *(Higher Scores Indicate Thinner Boundaries)*								
	3	187.7	(1.5)	165	246.9	(39.9)	6.5579	166	.01
	Myers-Briggs Type Indicator *(Scores below 100 indicate sensing scores above 100 indicate intuition)**								
Total Score	15	88.3	(29.0)	738	103.4	(29.9)	3.7513	751	.05
	Learning Styles Profile								
Sensing-Intuition	3	3.0	(0)	199	4.3	(1.1)	4.3550	200	.04
Simultaneous Visual Processing									

•Excerpted from Ehrman (1994b), Table 1.

Table 5

Results of ANOVAs for Strongest Students Compared with All Others

	Strongest Students			Non-Strongest Students					
Category	N	Mean	(SD)	N	Mean	(SD)	F	df	Sig.
From the Biographic Data									
Education Level	27	4.1	(1.2)	645	3.3	(1.1)	13.3136	670	.0003
No Previous Languages	26	2.3	(0.7)	637	1.6	(1.0)	10.0750	661	.002
High Previous Speak Score	21	3.2	(1.0)	331	2.2	(1.1)	17.0908	350	.0000
High Previous Read Score	21	3.5	(0.9)	325	2.3	(1.1)	23.0790	344	.0000
Age	27	33.7	(7.9)	584	39.4	(9.3)	9.6396	609	.002
Modern Language Aptitude Test									
Part 1	14	40.5	(4.9)	224	35.0	(9.7)	4.4395	236	.04
Part 2	14	27.2	(2.8)	224	24.3	(4.7)	5.2765	236	.02
Part 3	14	32.8	(7.0)	224	26.9	(10.2)	4.5701	236	.03
Total Score	14	151.2	(13.8)	224	132.2	(29.6)	5.7291	236	.02
Index Score	19	68.2	(5.9)	269	60.9	(11.2)	7.8286	286	.006
Hartman Boundary Questionnaire *(Higher Scores Indicate Thinner Boundaries)*									
Tolerate Lack of Orderliness	153	23.7	(5.0)	9	19.5	(6.3)	93.8905	160	.05
Myers-Briggs Type Indicator (TDI) *(Scores below 5 indicate sensing or judging poles; scores above 5 indicate intuition or perceiving poles)*									
Realistic-Imaginative	21	7.2	(2.9)	575	5.7	(3.2)	4.5036	594	.03
Methodical-Emergent	21	4.6	(2.8)	575	3.2	(2.5)	6.0914	594	.01

•Excerpted from Ehrman (1994b), Table 2.

Reading. For reading, Parts 3 and 5 and the Total and Index Scores best differentiate the weakest students. The strongest are differentiated clearly by all MLAT parts except Part 4, with the Index Score providing the clearest distinction.

Multiple Regression

An exploratory stepwise multiple regression analysis for end-of-training speaking and reading examined the effects of age, education level, number of previous languages studied, highest previous speaking and reading ratings, a general motivation rating, two self-efficacy ratings (self-rated aptitude and expectation of success in this course), two anxiety ratings (for the course in general and about speaking in class), and the MLAT Index Score.

For speaking, the analysis yielded a multiple R of .40, R Square of .16, with two predictors in the equation: the MLAT Index Score (Beta .32, T = 3.293 p = .0014) and Highest Previous Reading Score (Beta .21, T = 2.208, p = .0297).

For reading, the analysis yielded a multiple R of .37, R Square of .14, with the same two predictors in the equation: the MLAT Index Score (Beta .27, T = 2.798, p = .0063) and Highest Previous Reading Score (Beta .22, T = 2.266, p = .0258).

Results Related to Diagnosis and Student Counseling

In this section, both quantitative and qualitative findings are described, as part of an ongoing effort to build learner profiles that can be used by teachers, teacher trainers, program managers, and even students themselves to enhance student learning. The quantitative results contribute to a fuller picture of the kinds of students who are advantaged and disadvantaged in full-time intensive and largely communicative language training, by adding personality factors to more cognitive abilities. The qualitative material is very exploratory, but it has been promising enough to merit description here so that others can use and test the emerging patterns. It is also included here because it provides more information on what the MLAT may actually be measuring, and because it sheds more light on the complexity of the apparently simple factor-analysis-based MLAT parts.

Relationships with Other Individual Difference Variables

There are other variables than the MLAT that are useful in the building of an individual learner profile that can be used for diagnosis and counseling (the utility of these for prediction is more directly addressed in Ehrman, 1993, 1994a, b; 1995b, 1996, Ehrman & Oxford, 1995; Oxford & Ehrman, 1995). These variables bear interesting relationships to the MLAT. Correlations of at least .30 between the MLAT Index Score and/or Total Score and other instruments used in the larger study are presented in Table 6. The correlations suggest the relationships described below.

Table 6

MLAT Index or Total Score Correlations with Other Variables

Variable	Lang.Category Group	rho	Correlate	N
Number of Previous Languages	All	.40**	Index	245
HBQ Prefer Blurred Edges	Cat. 1	.51*	Total	25
HBQ Prefer Low Neatness	Cat. 2	.47	Total	25
HBQ Thin External Boundaries	All	.32**	Total	102
HBQ Total Score (Thin)	All	.30**	Index	110
MBTI/TDI Intellectual (N)	Cat. 1	.45*	Index	96
MBTI/TDI Intellectual (N)	Cat. 2-3	.35**	Index	103
MBTI Intuition	Cat. 1	.34**	Total	93
MBTI Imaginative (N)	Cat. 1	.34**	Index	96
MBTI Introversion	Cat. 1	.30*	Total	93
LSP Simultaneous Processing	Cat. 1	.45	Index	24
LSP Sequential Processing	Cat. 1	.43	Index	24

All the above correlations are significant at least at the .05 level; * indicates the .01 level, ** indicates the .001 level. HBQ: Hartmann Boundary Questionaire, MBTI, LSP: Learning Style Profile. "Imaginative" and "Intellectual" represent intuitive poles of the MBTI/TDI Realistic-Imaginative and Pragmatic-Intellectual subscales for the sensing-intuition main scale.

Those who have scored high on the MLAT tend to have studied languages previously and often prefer an "intuitive" approach to taking in information on the MBTI. MBTI intuition indicates preferences for the abstract over the concrete, search for meaning, a preference for the "big picture" rather than details, and the speculative over the strictly experiential (Myers & McCaulley, 1985). They describe themselves as having relatively thin ego boundaries, especially with respect to such matters as dislike for too much neatness, order, and clear-cut separations among visual images. Thin ego boundaries, correlated with MBTI intuition, indicate receptivity to a wide range of experience, both internal and external, and a willingness to blur categories. This concept is used to operationalize a model of tolerance of ambiguity (Ehrman, 1993, 1996, 1998). High-MLAT students also are often more skilled at simultaneous and sequential visual processing on the Learning Style Profile (Keefe, Monk, et al., 1989).

The analyses of variance in the study of extremes support these findings for extremely strong and weak students and add as an advantage a preference for a flexible approach shown in the perceiving pole of one of the MBTI/TDI JP subscales, methodical versus emergent. (This subscale of the TDI scoring of the long MBTI opposes a desire to know in advance what will happen to a preference to let events "emerge" and cope with them as they come up; the strongest students indicated a preference for an emergent approach.) Some of the results from the study of extremes are displayed in Tables 4 and 5.

The MLAT and Learning Activities

A recent correlation study showed interesting relationships between the MLAT and a set of activities that students rated for perceived utility both before starting training and at the end of training (Ehrman, 1995). The results led to hypotheses about the meanings of high and low subscale scores in a given student's profile that are proving useful in student counseling, so that the MLAT can be used for more than gatekeeping.

The correlations were similar for both pre- and post-testing. Though the correlations were generally low (mostly 20s and some in the 30s), there seemed to be suggestive patterns in them when subjected to a content analysis. Findings described below were based on the content analysis of those items with which the MLAT was correlated (Table 7) and on correlations of MLAT scales with variables.

Table 7
MLAT Index or Subscale Score Correlations with Items from Student Learning Activities Questionnaire

Pre-training questionnaire N=127; Post-training questionnaires N=various arranged from highest to lowest correlations

Items	Pre/Post Training	r	p	N
Index Score				
Positive				
Rates own ability as learner as good	Pre	.43	.000	127
Risk-taking is encouraged	Post	.40	.000	84
Forced to use what one knows to convey meaning	Post	.38	.000	86
Thinks has ability relative to FSI learners	Pre	.38	.000	127
Over-the-head reading is useful	Pre	.32	.000	127
Expectations of success in the course	Pre	.30	.001	127
Listen only to native speakers of target lang.	Post	.30	.000	79
Over-the-head listening is useful	Pre	.29	.001	127
Tries saying things one does not know the words for	Post	.26	.016	83
Study at home (vs. lab, library, other places)	Post	.25	.041	69
Role-plays, simulations, skits are useful	Post	.22	.050	79
Making presentations in the target lang. helps	Post	.22	.050	80
Negative				
Amt. of time spent making up tests for self	Post	-.55	.015	19
Amt. of time spent labeling pictures	Post	-.47	.041	19
Lang. Lab. (listening to tapes) useful	Post	-.33	.030	43
Amt. of time spent in lang. lab. listening to tapes	Post	-.33	.028	46
Go step-by-step so will not become confused	Pre	-.28	.002	127
Master one thing before going on to another	Post	-.26	.019	83

Part 1: Number Learning

Positive

Risk-taking is encouraged	Post	.37	.001	74
Forced to use what one knows to convey meaning	Post	.25	.029	76
Likes to help design the program as it goes along	Pre	.20	.033	127

Negative

Amt. of time spent making up tests for self	Post	-.45	.051	19
Amt. of time spent studying notes	Post	-.31	.013	62
Master one thing before going on to another	Post	-.26	.028	74
Teacher's role is to make and execute a plan	Post	-.23	.050	74

Part 2: Phonetic Script

Positive

Tries saying things one does not know the words for	Post	.32	.005	74
Risk-taking is encouraged	Post	.30	.011	74
Forced to use what one knows to convey meaning	Post	.29	.011	76
Study at home (vs. lab, library, other places)	Post	.28	.031	62
Use authentic magazines and newspapers	Post	.27	.042	56
Listen only to native speakers of target lang.	Post	.27	.025	70
Finds pattern drilling useful	Post	.27	.022	72
Observing native speakers interacting in real conversation	Post	.26	.025	75
Thinks has ability relative to FSI learners	Pre	.25	.022	127
Wants to have a syllabus	Pre	.23	.001	127
Role-plays, simulations, skits are useful	Post	.23	.050	71
Rates own ability as learner as good	Pre	.21	.021	127
Prefers grammar explanations in the target lang.	Pre	.21	.021	127
Over-the-head reading is useful	Pre	.20	.036	127

Negative

Amt. of time spent labeling pictures	Post	-.50	.029	19
Pronunciation not corrected unless unintelligible	Post	-.26	.028	73
Group study is part of the program	Post	-.25	.040	63
Go step-by-step not to confuse	Pre	-.21	025	127

Part 3: Spelling Clues

Positive

Using target language informally outside class (e.g., lunch)	Post	.37	.050	29
Listen only to native speakers of target lang.	Post	.28	.021	70
Oral transformation drills are useful	Pre	.27	.003	127
Thinks has ability relative to FSI learners	Pre	.25	.021	127
Correction focuses on formal features	Post	.24	.046	72
Rates own ability as learner as good	Pre	.22	.016	127
Listen to others speaking before trying to speak	Post	.22	.050	75

Negative

Making up tests for self is useful	Post	-.62	.011	16
Class content is related primarily to real life needs	Post	-.37	.001	73
Amt. of time in lang. lab. with tapes	Post	-.29	.040	43

Part 4: Words in Sentences

Positive

Listen only to native speakers of target lang.	Post	.44	.000	70
Using target language informally outside class (e.g., lunch)	Post	.36	.029	36
Forced to use what one knows to convey meaning	Post	.33	.003	76
Over-the-head listening is useful	Pre	.28	.002	127
Over-the-head reading is useful	Pre	.27	.005	127
Discover grammar patterns for self	Post	.26	.027	73
Risk-taking is encouraged	Post	.26	.026	74
Read without a dictionnary	Post	.25	.035	70

Correction focuses on formal features	Post	.24	.043	72
Discover grammar patterns for self	Pre	.20	.032	127
Negative				
Amt. of time in lang. lab with tapes	Post	-.46	.002	43
Amt. of time spent labeling pictures	Post	-.36	.029	19
Lang. lab. (listening to tapes) is useful	Post	-.36	.027	39
Master grammar before using in communication	Post	-.32	.006	74
Go step-by-step so will not become confused	Pre	-.29	.002	127
Studying one's notes is useful	Post	-.29	.036	16
Group study is part of the program	Post	-.29	.022	63
Reading is limited to what student already knows	Post	-.25	.041	70
Teacher's role is to make and execute plan	Post	-.25	.029	74
Field trips are useful	Pre	-.24	.009	127
Teachers are flexible and alter plans as needed	Post	-.24	.040	74
Master one thing before going on to the next	Pre	-.20	.032	127
Part 5: Paired Associates				
Positive				
Forced to use what one knows to convey meaning	Post	.29	.010	76
Listen to others speaking before trying to speak	Post	.25	.029	75
Over-the-head reading is useful	Pre	.23	.015	127
Negative				
Making up tests for self is useful	Post	-.65	.003	19
Recording self to listen to is useful	Post	-.49	.048	17
Proportion of time spent studying with other students	Post	-.41	.043	25
Lang. lab. (listening to tapes) is useful	Post	-.32	.048	39

In summary, high MLAT Index and the part scores correlate with items that are interpreted as reflecting self-confidence as a language learner and tolerance of ambiguity (low-structure activities and input). Very limited approaches, such as not moving on until something is mastered or strictly step-by-step learning, tend to correlate negatively with the MLAT.

The Index and Parts 2, 3, 4, and 5 are correlated with items suggesting acceptance of/preference for use of authentic material for reading and listening and authentic conversation.

Parts 3 and 4 are correlated with items suggesting endorsement of learning activities that reflect an analytic structured approach. This effect was slightly stronger for Part 3; students who rejected a "touchy feely" approach on one item (the only such item) also tended to be high scorers on Part 3.

In contrast, the Index and a strong peak score on Part 2 may suggest a more experiential approach, with many unstructured activities like role-plays that are supported by a syllabus and drilling so that learners do not have to do a lot of analysis on their own.

High scores on Part 3 and the Index appear to be related to comfort with unstructured auditory input.

Students who endorsed activities interpreted as indicating a preference for discovery learning tended to do well on the Index and Part 4.

Interpreting Part-Score Profiles

The above patterns suggested possible uses for the MLAT profile in student counseling, where they currently are being tested. Some profiles that these data suggest are outlined below.

1. All parts high (a very high Index will usually represent this kind of profile):

- has done well on all the parts
- self-confident as a learner
- responds well to activities that require tolerance of ambiguity
- likes relatively unstructured learning
- enjoys and even prefers authentic input.

A related analysis found a relationship between endorsement of relatively unstructured, ambiguous, authentic activities and higher end-of-training scores (Ehrman, 1995a).

2. A more uneven profile in which Parts 3 (especially) and 4 are high:

- analytic learner, perhaps field independent
- likes a program with a clear plan (not the same as a restrictively sequential program)
- usually has good knowledge of English vocabulary and grammar.

3. An uneven profile in which Part 2 is highest, together with a strong Index (most other parts above average), may indicate a student who likes experiential, hands-on, participatory learning and learns best from material in context.

4. An uneven profile in which Parts 2 and 4 are relatively high, together with a strong Index, may suggest a student who likes to take control of his or her own learning sequence and can use both analytic and global learning strategies comfortably.

5. When either Part 1 or Part 5 is the highest of the part scores, there so far seems to be little that is distinctive, though interviews are suggesting that low scores on Part 5 indicate either poor mnemonic skills or weak metacognitive strategies, or both.

6. All parts low (a very low Index will usually represent this kind of profile):

- has done poorly on all the parts
- often lacks self-confidence as a learner and subject to anxiety because of slow progress
- likely to be overwhelmed by unstructured and uncontrolled input
- will need a great deal of scaffolding for longer than most other students
- likely to progress slowly.

Overall Total score on the MLAT or the Index gives a useful crude measure when it is either very low or very high: a very low Total or Index score indicates weakness in all the factors; a very high score suggests strength in all the factors. When the Index falls in the middle

range—roughly within a standard deviation of the mean—it becomes much more important to examine the "scatter" of the part scores.

Using Part Scores With Students

The student counseling activity uses the variations in part scores to initiate interpretations that are raised with the student to examine how he or she learns. Interpretation usually requires an interview of the student. Responses by students to the question "What happened when you were doing this part?" provide useful information about the skills tested in each part. Each of the MLAT factors probably represents a set of abilities. For example, Part 3 has proved particularly fruitful in the diagnostic process with students. Among the possible task requirements of this item are gestalt processing of the whole word; sound-symbol processing; rapid hypothesis testing of sound-symbol possibilities; shift in mental set; and semantic evaluation.

These task requirement possibilities are represented as student performance in the following six cases of poor outcome on Part 3, each of which is followed by implications for the classroom. The cases represent composites of responses actually received to the query about what happened while students were completing this sub-test. (Many examples of real cases with specific score profiles are in Ehrman, 1996.)

1. One student might have done poorly on Part 3 because of difficulty with the kinds of analytic activities often described as "field independent." This student is likely to have difficulty with induction of rules and patterns and with grammar-oriented activities that have little context. Students of this sort usually find more contextual learning helpful.

2. Another might do poorly on the same part because of a weak English vocabulary (among the possible causal factors: poor education, low intelligence). This student, if a native speaker of English,[7] may have difficulty with vocabulary learning (among other things) because a lack of concepts and background knowledge. The classroom may have to include activities to help this student build content background as well as language.

3. A third one experiences difficulties reorganizing schemata or with gestalt processing or shifting mental set. Part 3 makes considerable demands on a person's ability to shift mental set. Such a student may be more comfortable with relatively predictable activities and less so with open-ended ones and may need assistance in building skills for coping with the unfamiliar or unexpected.

4. Yet another student might have a *phonetic coding difficulty* of the sort described by Sparks, Ganschow et al. (1995), that is, working with sound-symbol relationships. He or she is likely to have corresponding low scores in Parts 1 and 2, which also require decoding of sounds. Such a student is likely to be handicapped in both speaking and reading and will need more time to absorb material. Kinesthetic input such as learning with realia, drawing, and acting things out is likely to help this student.

5. Links among extraversion, desire for language use outside the classroom, and MLAT Part 3 suggest a student with a *distractibility* factor. That is, a strongly extraverted student who is drawn to interpersonal interactions might not be as adept at the kind of focus that the puzzle solving aspect of Part 3 entails as one who tunes out the world more readily. Study strategies, including frequent breaks and setting up conditions to maximize concentration, might help a student who has difficulty concentrating.

6. Finally, a person who is reminded by Part 3 items of crossword puzzles and dislikes them has had an *affective* reaction which interferes with ability to use cognitive resources. Alternatives to "puzzle-solving" activities would probably help this sixth student, or perhaps cooperative learning when puzzle-like activities are part of the curriculum. The teacher would need to be alert to the affective impact of these activities.

Interpretation of a student's profile is made more complex by factors that can affect any or all of the parts of the test. In some cases, a low score on Part 3 (or any other part) may be the result of a mechanical error, such as marking in the wrong row of the answer sheet. Sometimes a student will say that he or she did not understand the instructions for a given part (this response raises questions about attention, motivation, or test-taking strategies). Some students ascribe low scores to fatigue, which is plausible especially for the later parts. Interpretation is further complicated by the fact that a student might suffer from several of these difficulties at once.

Discussion

Summary

Despite the effects of restricted range, skewed distribution, and

relatively limited ceiling (because of negative skew for this high-end sample), the MLAT remains the best predictor of the variables examined. In general, the Index Score is the most useful of the MLAT variables as a predictor (strong in all cases, and with highest correlation coefficients). Of the part scores, Part 3 is the strongest predictor. Part 3, with its dependence on knowledge of English vocabulary as well as ability to solve puzzles, may also be an indirect indicator of general intelligence. This would apply to both fluid ability, because of the cognitive restructuring required by the task, and to crystallized ability (vocabulary), and "g" or general intelligence, since general vocabulary is also considered to be the single best stand-in for overall intelligence (Anastasi, 1988, Wesche, Edwards, & Wells, 1982).

Is the MLAT more suitable for Western European languages than for non-Western languages? The question remains open. Correlations show stronger results for Category 1 languages than for 2, 3, and 4 languages. On the other hand, the substantial preselection of students suggested by the very skewed distribution and the restriction of range in the sample may account for this finding as much as appropriateness of the MLAT for non-European languages. Furthermore, the fact that the correlations for Category 4 language outcomes are actually better than those for Category 3 languages, despite substantial truncation of range, might suggest that the MLAT is actually a fairly strong predictor for these languages. (The higher correlations might also be related to the much smaller numbers for Category 4 languages.) We cannot test either hypothesis on the FSI language-student population as long as they are pre-selected using the MLAT.

Of the extended set of variables in the research project (including learning strategies, cognitive styles, motivation, anxiety, and personality variables), the MLAT Index Score also continues to be the strongest predictor, both in the correlation coefficients and ANOVAs of extremely weak and strong students. It is especially powerful as a selector of extremes.

In addition to the relatively crude information provided by the Index score that may help in selection for training, the part-score profile shows promise as a way to better target classroom interventions and advice to students about appropriate learning strategies to develop. Strong performance on the MLAT appears to be related to personality variables that indicate high tolerance for ambiguity and the ability to reconceptualize input (e.g., reanalyze, arrange hierarchically, find abstractions that reconcile apparent contradictions).

Is the MLAT Passe in an Age of Communicative Teaching?

The MLAT has been criticized by many as rating aptitude only for audio-lingual training, which was in vogue when the MLAT was developed. However, the MLAT correlations remain about the same, although the teaching methodology has changed considerably (most FSI courses now have a substantial communicative component, and some are almost wholly communicative). Why is this so? The following are some possibilities.

1. Perhaps the MLAT is really multidimensional, and a different set of dimensions applies to different methodology.

2. Perhaps the operative factor is really some form of coping with ambiguity or coping with the unfamiliar.

3. Possibly, it is the "g" (general intelligence)-factor that is operative for FSI students. (Sasaki (1993) found a general cognition factor, which she describes as similar to "g," to account for 42% of the variance among Japanese college students studying English as a foreign language.)

4. The very nature of classroom training may make a difference. Although FSI classroom training requires the ability to cope with communicative activities and access global and inferential learning, it also makes heavy demands on analytic skills. These may become increasingly important at higher proficiency levels. This fact may be why Parts 3 and 4 together are the most predictive of extremes in achievement, together with the Index, which is more associated with predilection for the more open-ended learning that is also necessary for achieving high proficiency levels in FSI classrooms. The study of ego boundaries using the Hartmann Boundary Questionnaire (Ehrman, 1993) found a similar construct, "tolerance of ambiguity," to be essential to effective classroom learning at FSI. In this study, thin ego boundaries that let a student take in new data were not enough alone—students had to impose some sort of mental structure on their intake and at the same time stay open to the fact that their structures were hypothetical. Investigation now under way is examining the applicability of the field independence construct to these findings, further information on which is to be found in Ehrman (1996, 1997).

The Aptitude Concept

Expanding the aptitude concept is one of the subjects of an ongoing investigation of individual differences in language learning.

The subject is discussed in greater detail in Ehrman, 1994b, 1995b, 1996.

Among the outcomes of the study is evidence for an expanded definition of aptitude that includes both cognitive aptitude (measured specifically for languages by the MLAT and more generally by cognitive aptitude tests) and personality factors that predispose a learner to cope with ambiguity and apparent chaos. These become especially important in the relatively unstructured learning setting of communicative teaching approaches. A nexus is emerging of the following characteristics that seem to be related to success in the demanding intensive FSI classroom:

- cognitive aptitude (may include ability to cope with the un familiar)
- non-linear, discovery learning
- orientation to meaning over form
- ability to cope with surprises (linguistic and pedagogical)
- openness to input and tolerance of ambiguity
- ability to sort input, analyze as appropriate, and organize into mental structures.

The last is almost certainly related in some way to the much-studied field independence construct (e.g., Brown, 1994; Chapelle & Green, 1992; Ehrman, 1996, 1997; Witkin & Goodenough, 1981). It may be that the MLAT provides a way to measure field independence through verbal activities, in contrast to the usual tests of ability to disembed geometric figures (e.g., Witkin, 1969). Such a measure might improve the value of the field independence construct for language learning.

Absence of the above-listed characteristics appears to disadvantage FSI learners, perhaps more than the presence of these variables advantages those learners (Ehrman, 1994a, b, 1995b, 1996).

There seems to be a kind of aptitude-personality nexus that consists of cognitive flexibility, tolerance of ambiguity (including ability to impose structure on input), and ability to make use of non-linear learning strategies that are well suited to work with input in forms aimed at native speakers of the target language.

The MLAT is the most powerful of the predictive variables used, even in programs that are very different from those in vogue when it was designed. It may be that the ability to manage unfamiliar and contradictory input leads both to success in communicative classrooms and to high scores on the MLAT. The MLAT may gain its

relative power because it requires the examinee to cope with the unfamiliar on tasks that at least partially simulate language learning tasks. In contrast, personality inventories ask about general life preferences, and strategy inventories do not address how the strategies are used but only whether the student is aware of using them. "Faking good" is nearly impossible on the MLAT, and malingering is vanishingly rare at FSI.

Although the MLAT provides strong information about classroom language learning ability, it is supplemented by personality variables. The significant correlations between the MLAT and the personality measures, though not strong (between .21 and .33), are consistent across personality questionnaire and MLAT subscales (Ehrman 1993, 1994a, b, 1995b). In all cases, MLAT scores are linked with variables that suggest tolerance for ambiguity.[8]

The links between the MLAT and personality variables suggest a role for the disposition to use one's cognitive resources in ways that go beneath the surface and that establish elaborated knowledge structures. Those who are open to new material, can tolerate contradictions, establish hypotheses to be tested, focus on meaning, and find ways to link the new with previous knowledge structures seem to have an advantage in managing the complex demands of language and culture learning. The weakest students appear to be overwhelmed by the chaos they encounter; the strongest meet it head on, and may even embrace it to a degree.

As of now, the answer to the question "Is the MLAT passe?" is: Probably not, though it has much the same limitations as a sole *predictor* of learning success that it has always had. It is quite good, especially if viewed as an indicator of learning dispositions that will affect classroom performance, but it probably should not be more than one tool in a toolkit. Scatter analysis of the part scores is a promising use for placement, counseling, and remediation, particularly in the hands of an evaluator who treats the scores as signposts to interpretations to be tested, not as absolute predictors.

Limitations of This Study

The greatest limitation of this study, like all those from FSI, is the question of generalizability. Use of a sample drawn from a high-end, pre-selected population in itself restricts range, affects distributions, and strongly indicates the need for replication with samples more typical of what the usual reader of this publication works with. For the MLAT, unlike any of the other instruments in the larger study, the use of

the instrument itself to help pre-select the sample severely limits both the statistical normality of the sample and our ability to make inferences from the findings.

The impossibility of establishing a truly normal distribution of MLAT scores in this sample also means that the statistical tests that assume normal distributions and similar sample sizes are used in unconventional ways. The number of tests conducted increases the chance of type I errors (false positives), though the consistency of findings over a number of variables may reduce the likelihood of such error. For these reasons, the findings reported here must be considered suggestive, not conclusive.

Next Steps

There is much more to look at in these data in the course of trying to find out what the MLAT is good for and what are its limitations. Among these are to seek normally distributed samples on which to replicate this study, begin multiple regression and discriminant analysis to see if the MLAT is a better predictor in combination with other variables; and find out what has happened with subjects who return from overseas and are tested—are they improved, worse, the same?

On the qualitative front, continued investigation can seek to confirm the working hypotheses described above in the section on student counseling and systematize them for use by people other than researchers, so that the MLAT part scores can provide useful information about specific learning strengths and difficulties that can be used in curriculum design and interventions with individual students. Eventually, a quantitative study of the part-score profiles should be designed and undertaken.

Notes

[1]The remainder of the literature review owes much to a draft prepared by Frederick Jackson for an FSI roundtable at the Language Testing Research Colloquium in 1994 (Jackson, 1994).

[2]The MLAT Project is separate but overlaps with the Language Learning Profiles Project, especially because it uses the same data set.

[3]The Department of Defense uses a similar classification.

[4]Only three percent of students in this sample were studying

Category 2 languages, a number which is too small for most analyses. Category 2 and 3 languages are therefore combined.

[5]Although Appendix A lists possible Index Scores below 20, currently used scoring devices do not yield Index Scores below 20.

[6]The MLAT was standardized in part on an FSI sample. Although that sample, as a result of the times (late 1950s) was all male, no gender differences have appeared on the MLAT among present students on any sub-test of the MLAT or on its Total or standardized score.

[7]The MLAT is designed for use with native speakers of English. At FSI it is considered invalid for non-native speakers, though if one takes it and does well (Index greater than 50), such performance is considered a promising sign. Low scores, on the other hand, are ignored.

[8]A very recent study (Ehrman & Leaver, 1997), also shows a correlation of the MLAT with self-report of 'field sensitivity' (Index, r=.58, Part 2 .61, Part 3 .46, all at a p level of 0001). Field sensitivity, discussed at greater length in Ehrman (1996, 1997), is the tendency or ability to absorb language *osmotically* from the surrounding environment, in contrast to the field independent emphasis or decontextualizing the context.

References

Adams, M. (1984). The Modern Language Aptitude Test (Percentile Ranks). Unpublished FSI document. Arlington, VA: Foreign Service Institute.

Anastasi, A. (1988). *Psychological testing.* 6th ed. New York: Macmillan.

Brown, H. D. (1994). *Principles of language learning and teaching* (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall Regents.

Campbell, C. (1987). Survey of attitudes specific to the foreign language classroom. Unpublished manuscript.

Carroll, J. (1962). The prediction of success in intensive foreign language training. In Glaser, R. (Ed.), *Training research and education.* Pittsburgh, PA: University of Pittsburgh.

Carroll, J. (1990). Cognitive abilities and foreign language aptitude: Then and now. In T. Parry, & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs,

NJ: Prentice Hall.

Carroll, J. & Sapon, S. M. (1959). *Modern Language Aptitude Test*. New York: Psychological Corporation.

Chapelle, C., & Green, P. (1992). Field independence/dependence in second language acquisition research. *Language Learning, 42,* 47-83.

Clark, J. (1986). *A study of the comparability of speaking proficiency across three government language training agencies*. Washington, DC: Center for Applied Linguistics.

Demuth, K. A., & Smith, N.B. (1987). The foreign language requirement: An alternative program. *Foreign Language Annals, 20,* 67-77.

Ehrman, M. E. (1993). Ego boundaries revisited: Toward a model of personality and learning. In J. E. Alatis (Ed.). *Strategic interaction and language acquisition: Theory, practice, and research.* Washington, DC: Georgetown University Press.

Ehrman, M. E. (1994a). The Type Differentiation Indicator and adult language learning success. *Journal of Psychological Type, 30,* 10 - 29.

Ehrman, M. E. (1994b). Weakest and strongest learners in intensive language training: A study of extremes. In C. Klee (Ed.) *Faces in a crowd: Individual learners in multisection programs.* Boston MA: Heinle & Heinle.

Ehrman, M. E. (1995a). Correlations between MLAT scales and preferred student learning activities. Unpublished data.

Ehrman, M. E. (1995b). Personality, language learning aptitude, and program structure. J. Alatis, (Ed.), *Linguistics and the education of second language teachers: Ethnolinguistic, psycholinguistic, and sociolinguistic aspects* (pp. 328-345). Washington, DC: Georgetown University Press.

Ehrman, M. E. (1996). *Understanding second language learning difficulties*. Thousand Oaks, CA: Sage Publications.

Ehrman, M. E. (1997). Field independence and field sensitivity. In Reid, J. (Ed.), *Understanding Learning Styles in the Second Language Classroom.* Englewood Cliffs, NJ: Prentice Hall Regents.

Ehrman, M. E. (1998). Ego boundaries and tolerance of ambiguity in second language learning. In J. Arnold (Ed.), *Affect in language learning*. New York: Cambridge.

Ehrman, M. E., & Christensen, L. (1994). *Language learner motivation and learning strategies questionnaire*. Unpublished

manuscript.
Ehrman, M. E., & Jackson, F. H. (1992). Classroom activities survey. Unpublished manuscript.
Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 1,* 67-89.
Ehrman, M. E., & Leaver, B. L. (1997, March). Sorting out global and analytic functions in second language learning. American Association for Applied Linguistics Annual Meeting, Orlando, FL.
Hart-Gonzalez, L. H., & Ehrman, M. E. (1992). Study Activities Questionnaire. Unpublished manuscript.
Hartmann, E. (1991). *Boundaries in the mind: A new psychology of personality*. New York: Basic Books.
Jackson, F. H. (1994). Language aptitude. Draft talking points prepared for FSI roundtable on the Modern Language Aptitude Test, annual meeting of the Language Testing Research Colloquium,Washington, DC.
Keefe, J. W., & Monk, J. S., (with Letteri, C.A., Languis, M., & Dunn, R). (1989). *Learning Style Profile*. Reston, VA: National Association of Secondary School Principals.
Lefrancois, J. & Sibiga, T. C. (May, 1986). Use of the Modern Language Aptitude Test (MLAT) as a diagnostic tool. Unpublished paper, no source. [Ontario Institute for Studies in Education?]
Lett, J. A., & O'Mara, F. E. (1990). Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defense Language Institute Foreign Language Center. In Parry, T., & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 222-260.) Englewood Cliffs, NJ: Prentice Hall.
Lett, J. A., & Thain, J. (1994). *The Defense Language Aptitude Battery: What is it and how well does it work?* Paper presented at the Language Aptitude Invitational Symposium, Arlington, VA.
Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists.
Norusis, M. J. (1992). *SPSS for Windows 5.0.1*. Chicago: SPSS.
Oxford, R. L., & Ehrman, M. E. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System, 23,* 359-386.
Parry, T. S., & Child, J. R (1990). Preliminary investigation of the rela-

tionship between VORD, MLAT, and language proficiency. In Parry, T. & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 30-66). Englewood Cliffs, NJ: Prentice Hall.

Parry, T., & Stansfield, C. W. (1990). Introduction. In T. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp.1-10). Englewood Cliffs, NJ: Prentice Hall.

Parry, T., & Stansfield, C. W. (Eds.). (1990). *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall.

Petersen, C. R., & Al-Haik, A. R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and psychological measurement, 6,* 369-380.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt, Brace, Jovanovich.

Pimsleur, P. (1968, August). Aptitude testing. *Language Learning*. Special Issue, *3*, 73-78.

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning, 43,* 313-344.

Saunders, D. (1989). *Type Differentiation Indicator Manual: A scoring system for Form J of the Myers-Briggs Type Indicator.* Palo Alto, CA: Consulting Psychologists.

Skehan, P. (1989). *Individual differences in second language learning.* London: Edward Arnold.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*(2), 275-278.

Sparks, R. L., Ganschow, L., and Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native language and foreign language learning. *Journal of Educational Psychology, 87,* 638-655.

Spolsky, B. (1995). *Measured words: The development of objective language testing*. New York: Oxford.

Thain, J. (1992). DLAB II prototype development: Status report and CY92 plan. Technical report. Monterey, CA: Defense Language Institute Foreign Language Center.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.

Wesche, M., Edwards, H., & Wells, W. (1982). Foreign language aptitude and intelligence. *Applied Psycholinguistics,* 127-140.

Wilds, C. P. (1965). MLAT Index Scores: Foreign Service Institute Curve-1965. Unpublished table. Arlington, VA: Foreign Service Institute.

Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: Essence and origins: Field dependence and field independence*. New York: International Universities.

Witkin, H. A. (1969) *Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists.

Appendix A

MLAT Subscales

Part 1

Number Learning: This sub-test requires the examinee to learn four morphemes and interpret them in combinations that form numbers; it is entirely orally delivered. The sub-test is described in the Manual (Carroll & Sapon, 1959) as measuring part of memory and "auditory alertness" which play a part in auditory comprehension (showing how well one understands what one hears) of a foreign language.

Part 2

Phonetic Script: This sub-test requires the examinee to select a written equivalent (in Trager-Smith phonemic transcription) for an orally delivered stimulus. The MLAT Manual describes the sub-test as dealing with the ability to associate a sound with a particular symbol, as well as how well one can remember speech sounds. In addition, the sub-test is described as tending to correlate with the ability to mimic speech sounds and sound combinations in a foreign language.

Part 3

Spelling Clues: In this entirely written sub-test, an English word is presented in a very non-standard spelling. The examinee must select structure and thus expected to provide information about the ability to handle grammar in a foreign language. No grammatical terminology is used, so scores do not depend on specific memory for grammatical terms.

Part 4

Words in Sentences: The stimulus is a sentence with a word or phrase highlighted. The examinee must indicate which part of another sentence matches the designated part. The sub-test is entirely in writing. It is described as dealing with the examinee's sensitivity to grammatical with their English equivalents and given some time to learn them. The words are then tested. This sub-test is said to measure the examinee's ability to memorize by rote—a useful skill in learning new vocabulary in a foreign language.

Part 5

Paired Associates: The examinee is presented with 24 foreign words he correct synonym. Vocabulary items are progressively more difficult, though the most difficult is probably within the repertoire of a college graduate. According to the Manual, scores on this part depend largely on how extensive a student's English vocabulary is. As in Part 2, it measures the ability to make sound-symbol associations but to a lesser degree.

Raw Score Total: Total of all five subscales.

Index Score: Originally a scaled (T) score used at FSI that is based on the Total. The original mean was 50, with a standard deviation of 10. These norms are now out of date; the Index is now simply a conversion of the raw Total into a scale ranging between 20 and 80. Local norms using the Index have not been formally established because the Index score using the original norms is deeply embedded in the agency's personnel system.

Appendix B

Conversion Table for MLAT Raw Total and Index Scores

Raw Total	Index	Raw Total	Index	Raw Total	Index
0-9	15	67-68	37	125-127	59
10-12	16	69-71	38	128-129	60
13-15	17	72-74	39	130-132	61
16-18	18	75-76	40	133-135	62
19-21	19	77-79	41	136-137	63
22-23	20	80-82	42	138-140	64
24-26	21	83-84	43	141-143	65
27-29	22	85-87	44	144-145	66
30-31	23	88-90	45	146-148	67
32-34	24	91-92	46	149-150	68
35-37	25	93-95	47	151-153	69
38-39	26	96-97	48	154-156	70
40-42	27	98-100	49	157-158	71
43-44	28	101-103	50	159-161	72
45-47	29	104-105	51	162-164	73
48-50	30	106-108	52	165-166	74
51-52	31	109-111	53	167-169	75
53-55	32	112-113	54	170-172	76
56-58	33	114-116	55	173-174	77
59-60	34	117-119	56	175-177	78
61-63	35	120-121	57	178-180	79
64-66	36	122-124	58	181-182	80

Author

MADELINE EHRMAN, Foreign Service Institute U.S. Department of State, Mailing address: 1501 South Randolph St., Arlington, VA 22204-4012. Specializations: individual differences (learning styles and personality), interpersonal dynamics in second language education, and program leadership.

Applied Language Learning
1998, Vol. 9, Nos. 1 & 2, pp. 71-105

Factors in the Prediction of Achievement and Proficiency in a Foreign Language

Richard L. Sparks
College of Mount St. Joseph

James Javorsky
Purdue University

Jon Patton and **Leonore Ganschow**
Miami University

This study examines the results of a factor analysis of a battery of native language and foreign language (FL) aptitude measures used to predict FL proficiency. The study involved two groups of high school FL learners completing a second year of FL study, sixty 10th and 11th grade females attending a private, single sex, college preparatory high school and a coeducational population of thirty-six 10th grade students in a public school. The two groups were combined to perform the analysis. The latent structure of the test battery was of interest to determine if the components that emerged in previous factor analyses to predict FL grade would differ from the components that emerged in the present study to predict oral and written FL proficiency. Three components emerged from the principal components analysis and were identified as: Verbal Memory, Phonological Coding/Recoding, and Cognitive Speed Plus. Results showed that the three components received similar loadings in the analysis; together, the three components contributed 63.28% of the variance in overall FL proficiency.

© 1998, Richard L. Sparks, James Javorsky, Jon Patton and Leonore Ganschow

Recently, foreign language (FL) educators have developed guidelines to measure the extent to which students become proficient in the oral and written aspects of a FL. The American Council on the Teaching of Foreign Languages (ACTFL) defines proficiency as "what an individual can and cannot do (with a foreign language), regardless of where, when, or how the language has been learned or acquired" (ACTFL Proficiency Guidelines, 1989). In part, because FL proficiency tests require trained professionals and are time consuming, only a few studies have used these guidelines. These studies have shown that ratings on FL proficiency measures can be useful as criterion variables and are closely related to scores on norm-referenced tests of proficiency in similar language domains (see Ehrman & Oxford, 1995; Lett & O'Mara, 1990; Wilson, 1989; Wilson & Graves, 1995).

For over forty years, FL educators and researchers have investigated the best predictors of FL learning success. In the 1950s and 1960s, researchers developed FL aptitude tests. The underlying assumption of these tests was that FL aptitude is closely associated with the linguistic rule structures of language. Paul Pimsleur developed a measure of FL aptitude, the *Language Aptitude Battery* (LAB) (Pimsleur, 1966), that was closely associated with analysis of linguistic structures. The LAB is comprised of a Verbal Ability score (vocabulary, language analysis) and an Auditory Ability score (sound discrimination, sound-symbol association). John Carroll and Stanley Sapon published the Modern Language Aptitude Test (MLAT, Carroll & Sapon, 1959) based on Carroll's ground-breaking factor analytic studies, the results of which showed four independent variables to be important for FL learning: (1) phonetic coding; (2) grammatical sensitivity; (3) inductive language learning ability; and (4) rote memory (Carroll, 1962). In a comprehensive factor analytic study of cognitive abilities, Carroll (1993) proposed a three-stratum theory in which FL aptitude and FL proficiency are both distinct and unique factors. FL aptitude and FL proficiency appear in the first, or lowest, stratum of Carroll's hierarchical model of cognitive abilities. The second stratum of Carroll's theory encompasses eight unique abilities, two of which are called "crystallized intelligence" and "broad cognitive speed." Of importance to FL learning theory is that FL aptitude (including phonetic coding, grammatical sensitivity, spelling, and verbal (printed) language) is subsumed by "crystallized intelligence." In his Model of School Learning for the study of a FL, Carroll also proposed that time, i.e., "broad cognitive speed," is important for FL learning. (See Skehan, 1986, for a discussion of Carroll's Model of School Learning.)

In a recent study in which they examined the correlational and predictive validity of a new language aptitude test similar to the MLAT, Parry and Stansfield (1990) found that the MLAT was the best overall instrument for predicting language learning success in an adult population. Lett and O'Mara (1990) found that another FL aptitude test, the Defense Language Aptitude Battery (DLAB) (Petersen & Al-Haik, 1976), along with high cognitive ability were consistent predictors of success among learners in intensive FL courses in more difficult FLs. Although hypothesized to offer potential in enhancing success in FL learning, non-cognitive variables (e.g., attitude, motivation, language learning strategies, personality, cognitive style) were not predictive of FL learning success. FL aptitude tests were widely used in the 1960s and 1970s, but then generally fell out of favor with FL educators and researchers.

Other researchers have speculated that the concept of FL aptitude needs to be reexamined because "FL aptitude tests such as the MLAT do not take into account new insights revealed by cognitive psychologists into the human learning process in general, and the language learning process in particular" (Parry & Stansfield, 1990: p. 2). In the 1970s and 1980s, FL educators began to emphasize the role that affective variables might play in successful FL learning. For example, Gardner and his colleagues studied the role of attitudes and motivation in FL learning and hypothesized that these two variables were relatively independent of language aptitude (e.g., see Garnder 1985, 1990; Gardner & Lambert, 1972). Other researchers have investigated anxiety in FL learning and speculated that there is a type of anxiety specific to FL learning (e.g., Horwitz, Horwitz, & Cope, 1986; MacIntyre & Gardner, 1991, 1994). Ehrman (1990) has used measures of personality (e.g., Myers-Briggs Type Indicator, or MBTI) to determine their importance in predicting FL learning. Qualitative findings of her research show that personality measures can provide information about the compatibility of teaching methods and teaching styles with the learning styles of individual learners; quantitatively, however, personality instruments appear to be weak predictors of language learning success (Ehrman, 1990; Ehrman & Oxford, 1995). Oxford (1990a, b) suggests that language learning strategies and a learner's cognitive style might be responsible for success or failure in learning a FL. She and her colleagues suggest that good language learners use more efficient language learning strategies and that FL learners can be trained to use strategies to improve FL learning.

Some researchers have challenged theories that emphasize a role for affective factors and language learning strategies in predicting

FL success or failure. For example, Au (1988) has criticized Gardner's theories for failure to show a causal link between attitudes/motivation and FL learning. Oller (1981) has speculated that affective instruments may be unintentionally assessing language proficiency. Sparks and Ganschow (1991) support Oller's position, e.g., suggesting that most items on Horwitz's Foreign Language Classroom Anxiety Scale, a self-reported measure of anxiety, are related to expressive or receptive language and verbal memory skills, thus confounding the effects of anxiety and leading to speculation as to causal direction, i.e., which came first, the difficulties with language learning or the anxiety? They suggest that affective states, e.g., low motivation, high anxiety, are generally the consequence, rather than the cause, of good and poor FL learning. Tiedeman (1989) reviews cognitive styles research and finds that most measures of cognitive style are best interpreted as ability tests, not as measures of preferences for information processing. Skehan (1991) finds little evidence to suggest that style or strategy training improves FL learning and suggests that there is still the "worrying possibility that good [language] learners are ones for whom the use of effective strategies are possible, while for the poor language learners they are not" (p. 288). To the present authors' knowledge, previous research does not substantiate that affective variables, learning styles, or learning strategies are primary causal factors in FL learning, nor is research conclusive about whether these variables and others (e.g., beliefs about language learning, culture, gender) are predictive of FL proficiency.

Several years ago, Sparks, Ganschow, and Pohlman (1989) introduced into the learning disabilities and FL literature the hypothesis that the extent of one's proficiency in his/her native language is likely to have an impact on one's ability to learn a FL (see Sparks, 1995; Sparks & Ganschow, 1991, 1993a,b, 1995a). Their hypothesis is called the Linguistic Coding Differences Hypothesis (hereafter referred to as LCDH).[1] The term "linguistic coding" was coined initially by Vellutino and Scanlon (1986) to describe the language-based problems of children who had pronounced difficulties learning to read and spell their native language. Vellutino and Scanlon demonstrated that these children had difficulty primarily with the phonological/orthographic (sound and sound-symbol) and syntactic (grammar), but not the semantic (meaning) codes of language. In the LCDH, Sparks and his colleagues hypothesize that FL learning is built upon one's native language skills; that is, an individual's skill in the native language components — phonological/orthographic, syntactic, and semantic — serves as the foundation for successful FL learning. Further, they speculate that both native and

FL learning depend on basic language learning mechanisms and that problems with one language skill, e.g., semantics, are likely to have a negative effect on both the native and FL systems.

Native language researchers also have shown that students with difficulties in reading and writing often have overt or subtle problems with their oral language skills (e.g., Catts, 1986; Crain, 1989; Mann, Shankweiler, & Smith, 1984). In their research, Sparks and Ganschow have hypothesized that the majority of students with FL learning problems are likely to display problems with the phonological/orthographic (and sometimes, syntactic) codes of language.

Since the introduction of the LCDH, its authors have conducted a number of research studies at the secondary and postsecondary levels of education. Results of these studies have provided strong empirical support for the hypothesis (Ganschow & Sparks, 1991, 1995, 1996; Ganschow, et al., 1991, 1994; Javorsky, Sparks, & Ganschow, 1992; Sparks, Artzer, et al., 1998; Sparks & Ganschow, 1993a,c, 1995b, 1996; Sparks, Ganschow, Artzer, & Patton, 1997; Sparks, Ganschow, & Javorsky, 1993; Sparks, Ganschow, & Patton, 1995; Sparks, Ganschow, & Pohlman, 1989; Sparks, et al., 1992, 1992a, b, 1996, 1997, 1998).

Sparks and Ganschow have posited that the concept of FL aptitude and its relationship to native language skills holds the most potential for better understanding: (a) why (and how) some students learn a FL better than others; and (b) why language aptitude measures are good predictors of FL learning and success. They have advocated the position that language variables are likely to contribute the largest part of the variance in successful FL learning (Sparks & Ganschow, 1995a). They have encouraged FL researchers to consider what is known in native language learning as the Assumption of Specificity (AOS) as a guide to future research in FL learning. (See Hall & Humphreys, 1982, and Stanovich, 1988 for a detailed description of the AOS). The concept of the AOS proposes that students with a particular learning problem have a cognitive deficit (or difference) that is reasonably specific to the task in question; with regard to FL learning, the deficit (or difference) is likely to be related to language because FL learning is a language-based task. Sparks and Ganschow suggest that researchers who wish to determine the underlying cognitive factors that may be related to good or poor FL learning should begin by carefully specifying the particular performance differences whose nature and origin are of interest. In their view, the performance differences between good and poor FL learners should be language-related because it is highly probable that differences in FL learning are related to the learning of

language. (See, also, Sparks, 1995, and Sparks and Ganschow, 1993b, for a discussion of the AOS). Thus, the LCDH focuses specifically on language variables in FL learning, much like Carroll's model of FL aptitude.

One method to test the soundness of a hypothesis is factor analysis, which identifies variables that correlate highly and those that do not. Ganschow, Sparks, Javorsky, and Patton (1992) conducted a factor analysis of a test battery used to test the native language and FL aptitude skills of not-at-risk and at-risk FL learners at the secondary level. In the study, they found three separable components which they labeled Phonology/Syntax, Cognition/Semantics, and FL Aptitude. The three components contributed approximately 60% of the variance in students' end-of-year FL grades. The Phonology/Syntax component had the most significant loading in the analysis and was comprised of measures of spelling, word recognition, and pseudoword reading as well as grammar. They interpreted this component to be highly related to phonology/orthography because the written language measures used in the test battery were primarily composed of phonological/orthographic items (e.g., spelling, pseudoword reading).

Sparks, Ganschow, and Patton (1995) conducted a factor analysis of a test battery used to predict end-of-year FL grade. The test battery was similar to the one used in the aforementioned study with a different group of FL learners completing the first year of a high school FL course. In that study, they found three components that they labeled Phonology/Orthography, Meaning, and FL Aptitude/Metalinguistic. Again, the Phonology/Orthography component had the most significant loading in the analysis and the three components contributed approximately 60% of the variance in the end-of-year FL grades of the students. In these studies the results showed that both lower level processing, the sound and sound/symbol system of language represented by the Phonology/Orthography component (see Koda, 1992), and higher level processing, the semantic and metalinguistic components of language represented by the Meaning and FL Aptitude/Metalinguistic components, contributed to performance in the FL classroom.

To the authors' knowledge, no factor analytic studies have been conducted on a battery of testing measures used to predict FL proficiency.[2] Of interest in this present study was the determination of components that might emerge in a factor analysis of a test battery to predict oral and written FL proficiency and whether the factors that emerged would be different from factors identified in two prior factor analyses that were conducted to predict end-of-year FL grades (see Ganschow, et al., 1992; Sparks, Ganschow, & Patton, 1995). The purpose of the

present study, then, was to determine the latent structure underlying a battery of native language and FL aptitude tests, most of which had been administered to the students prior to the first year of FL study. Also included were two measures obtained during the course of the study: end of first-year grades in the FL course and a FL word recognition measure administered at the end of the students' second year of FL study. The authors were also interested in determining which of the factors that emerged from the factor analysis were predictive of performance in English courses, FL courses, FL word recognition, and FL proficiency.

Method

Participants

Participants were 96 students enrolled in the second year of a FL course in the tenth and eleventh grades. Sixty students were females attending a highly selective, single sex, college preparatory high school and 36 students (17 males, 19 females) were attending a large, middle class, suburban public high school.[3] The mean age of the 96 participants was 16 years, 1 month (age range = 15 years, 5 months to 17 years, 4 months). The participants were enrolled in three FLs (Spanish = 52, French = 27, German = 17). All of the students had participated in the authors' previous studies (Sparks, Ganschow, & Patton, 1995; Sparks, et al., 1998). The justification for combining the data from two schools was that: (a) both groups had been administered similar test batteries; (b) the combined data sets resulted in the inclusion of a larger population; and (c) the combined data sets increased the probability that a more consistent factor pattern would emerge.

Instruments for Factor Analysis

Three types of instruments — native language, FL aptitude, and FL word recognition — were used to predict various aspects of FL proficiency and achievement, end of first-year FL grades, and eighth grade English grade. The predictor measures and outcome variables are described below.

Predictor Measures.

Measures common to both private and public school populations were used in the factor analysis: (a) the Wide Range Achievement Test-Revised (WRAT-R): Spelling Subtest; (b) Peabody Picture Vocabulary Test-Revised (PPVT-R); (c) Woodcock Reading Mastery Test-Revised (WRMT-R) Basic Skills Cluster: Word Identification and Word Attack subtests; and (d) the five subtests from the Modern Language Aptitude Test (MLAT) (Parts I-V).

Three measures in the analysis differed in the two populations, but were conceptually similar and measured similar skills. First, a group achievement test had been administered by both schools (High School Placement Test (HSPT): Total Test at the private school, Iowa Tests of Basic Skills (ITBS): Total Test at the public school). Second, a group-administered reading comprehension test had also been administered by both schools (Nelson-Denny at the private school, ITBS: Reading Comprehension subtest at the public school). Third, two different author-designed phonemic awareness measures were administered at both schools (Phoneme Deletion in the private school, Pig Latin in the public school). The HSPT and ITBS: Total Test were combined to form a Group Achievement variable; the Nelson-Denny and ITBS: Reading Comprehension subtest were combined to form a Reading Comprehension variable; and the Phoneme Deletion and Pig Latin measures were combined to form a Phonemic Awareness variable. For the phonemic awareness tasks, Z scores were computed by standardizing both variables and were used in subsequent analyses.

A list and description of testing instruments is presented in Appendix A. Abbreviations for each variable are used in Appendix A and in all Tables.

Outcome Variables

The outcome variables concerned the participants' performance in their native language and in the FL. Four types of outcome measures were collected: (a) end-of-year grades in the first year of FL high school instruction; (b) end-of-year grades in eighth grade English courses; (c) overall proficiency in reading, writing, and speaking/listening to a FL; and (d) FL word recognition. End-of-year FL and English grades were obtained from the participants' school records. In both subjects, grades represented scores on homework, projects, in-class activities, and oral and written quizzes and tests. The FL instructors reported that grades were comprised of approximately 25% listening, 25% speaking, and

50% reading and writing activities.

The outcome variables in FL performance involved the measurement of the participants' FL proficiency (i.e., skill in reading, writing, speaking, and listening to the FL). Three university-level FL professors, who were formally trained to administer proficiency tests in their respective languages of Spanish, French, and German, designed the tests according to guidelines developed by the ACTFL (1986, 1989). The overall proficiency score (identified as FL Total Test) included measures of written (reading and writing) and oral (listening and speaking) abilities in their respective FLs. Scores on the written and oral measures were combined to obtain an overall proficiency score.

Two measures were used to assess the student's proficiency in reading a foreign language. The first was a fictitious letter written in Spanish from Claudia Rivera, a high school student in Argentina, to a family in the United States. Claudia was planning to spend a year as an exchange student with the family to whom she was writing. The letter contained information about her, her family in Argentina, and a series of five questions that she wished to have answered prior to her arrival in the United States. The student was given 15 minutes to read the letter and answer ten multiple-choice questions in English about the contents of the letter. The second measure of reading proficiency was a slightly more difficult passage. The student was given 15 minutes to read a brief article from Selecciones (i.e., Readers Digest in Spanish) entitled "Los Palos de Punta" and answer ten multiple choice questions in English about the contents of the article. The student could achieve a combined maximum score of 20 on this reading comprehension measure.

To assess writing in a foreign language, the student was given 15 minutes to write a letter to answer Claudia's letter, incorporating the answers to Claudia's five questions in the response. ACTFL Guidelines were used in assigning a holististic proficiency level (i.e., one score based on all the criterion statements in a specific level of the ACTFL Guidelines) on the writing test. After the holistic score was determined, the student's performance was further defined for quantitative purposes by assigning a score of 0-5 on each of the following writing skills: vocabulary, cultural appropriateness, structures, comprehensibility, and spelling (0 = no production, 1 = Novice-Low, 2 = Novice-Mid, 3 = Novice-High, 4 = Intermediate-Low, 5 = Intermediate-High). A score of 0 was included in the scoring because some students at this level of education may have been unable to produce any response in Spanish. A student could achieve a maximum score of 25 on the writing measure.

To measure the student's ability to listen to and speak a FL, a 10-15 minute oral interview following ACTFL Oral Proficiency Interview guidelines was conducted individually with each student using the four phases prescribed in the ACTFL Guidelines: warm-up, level check, probes, and wind-up (Omaggio, 1986). The entire interview was audiotaped for later scoring. Prior to the beginning of the oral interview, the tester explained to the student in English that, after she had had an opportunity to chat for a few moments in Spanish, she would be given a conversation card in English to help her begin the conversation (Spinelli, 1988). The interview proceeded as a friendly conversation in Spanish about topics which naturally emerged as the student responded to the conversation and the interviewer guided the conversation through the phases listed above. The ACTFL Proficiency Guidelines for determining proficiency levels were used in assigning a holistic score (i.e., one score based on the ACTFL Guidelines) on the oral interview. After the holistic score was determined, the student's performance was further defined for quantitative purposes by assigning a score of 0-5 on each of the following skills: pronunciation, vocabulary, grammar, comprehensibility, and listening comprehension. The scoring procedure for the listening/speaking test was the same as in the FL writing test. A student could achieve a maximum score of 25 on the listening/speaking measure.

A student's total test score (FL Total Test) was the combination of her scores on the reading comprehension, writing, and listening/speaking tests. A student could achieve a maximum score of 70 on FL Total Test.

The reliability of the three proficiency subtests (reading comprehension, writing, listening/speaking) and the total proficiency test (FL Total Test) were checked by a Cronbach's Alpha calculation. For reading comprehension, the Cronbach's Alpha was .73; for writing, .76; for listening/speaking, .97; and for FL Total Test, .87.[4]

To assess word recognition in Spanish, students were also asked to read a list of 20 words in Spanish, some of which they had never seen before. This served to assess directly phonological/orthographic skills in Spanish, and indirectly, pronunciation ability. Each word was chosen because it contained a letter or letter combination with a phonetic sound (e.g., the /a/ sound in the Spanish word casa is different from the /a/ sound in the English word cat) or a phonetic element that is different in Spanish than it is in English (e.g., in the Spanish word teléfono, the primary stress fall on the second syllable). The vowel sounds in Spanish, diphthongs, words with diacritical marks, and multisyllabic words were included within the target words. Refer to Appendix B for sample

lists of words in Spanish, French, and German. Due to an inadvertent error in communication, the German lists contained only half as many words as the Spanish and French lists, and the error was not noted until after completion of the testing. German students, however, were each administered two lists of ten words and, therefore, read the same number of words (20) as the Spanish and French students.

After the Spanish and French students' raw scores were obtained on the word recognition measure, raw scores were transformed into Z scores. For the German students who read two lists, raw scores were also transformed to Z scores. Each German student's score was the average of the standardized scores from the two lists. These standardized scores were used in all subsequent analyses.

Procedure

Eighth grade English grades and the results of the group achievement tests (HSPT and ITBS) were obtained by the authors from school records. The authors administered the native language measures and the FL aptitude test during the first quarter of the school year in which each participant was enrolled in a first-year FL course. The MLAT, WRAT-R: Spelling, and Nelson-Denny Reading Test were administered in groups. The PPVT-R, WRMT-R; Basic Skills Cluster, and the phoneme awareness measures (Pig Latin and Phoneme Deletion) were administered individually. Total test time was approximately two hours for the native language and FL aptitude measures.[5]

The oral (speaking/listening) and written (reading comprehension, writing) FL proficiency measures (FL Total Test) were administered at the end of the participants' second year of FL study. The FL reading comprehension test took ten minutes, the FL writing test took fifteen minutes, and the FL speaking/listening test took from ten to fifteen minutes to complete. The FL word recognition measure was administered after the FL speaking/listening test was completed. The interviewer randomly selected one of the five word lists in Spanish and French, each of which contained twenty words, and two of the five lists in German, each of which contained ten words. The word recognition measure took five minutes to administer. Both the FL Total Test and FL word recognition measure were administered at the end of the students' second year of FL study.

Data Analysis

To determine the latent structure underlying the testing instruments, a factor analysis with varimax rotation was performed on the data. Because the goal of the factor analysis was to derive factor scores for use in a multiple regression analysis, a varimax rotation was selected (Tabachnick & Fiddell, 1996). The purpose of varimax rotation, which is the most commonly used in social science research, is to simplify factors by maximizing the variance of the loadings within each factor across variables. The spread in loadings is maximized — loadings that are high after extraction become higher after rotation and loadings that are low become lower. Thus, interpreting a factor is easier because it is obvious which variable(s) correlates with it. To select the number of factors to analyze, we used a minimum eigen value, which represents a variance of 1.0. Because the variance that each standardized variable contributes to an initial factor analysis extraction is 1.0, a component with an eigen value of less than 1.0 is not as important from a variance perspective as an observed variable (Tabachnick & Fiddell, 1996).[6]

The analysis examined the relationship among and between observed variables and their relationship to a set of unobserved components or theoretical constructs. This relationship appears as a factor loading that has a range of 0 to 1.0. The higher the loading of an observed variable, the more important the variable is to the factor, or construct. From the resulting factors, predictor scores were derived. In order to examine the predictive relationship between the predictor scores and the outcome variables, a regression procedure was applied.

To examine the predictive relationship between the factors and the outcome variables, a standard multiple regression analysis was performed using Factor scores, derived from the factor analysis, as dependent variables, and FL grades, English grades, FL word recognition, and overall FL proficiency as independent variables. Statistical significance for this analysis was set at $p \leq .05$.

Results

Three factors based on the 12 tests and subtests emerged from the factor analysis and accounted for 63.28% of the variance of the model. Table 1 shows the three components, the testing measures, their component loadings, and the communality estimates for each of the testing measures. Factor 1 (accounting for 21.58% of the variance) was defined as a Verbal Memory dimension, Factor 2 (21.55%) as a

Table 1

Principal Components solution (with Varimax Rotation) and Communaliy Estimates on Test Battery in Combined Datasets

Test	Factor 1	Factor 2	Factor 3	Communality Estimates
MLAT I	**.71***	.19	.08	.55
MLAT V	**.71***	.22	.07	.55
PPVT-R	**.68***	.03	.25	.53
WRMT WATT	.17	**.81***	.32	.78
PHON AW	.21	**.77**	-.03	.64
WRAT-R SPELL	.13	**.69**	.47	.71
WRMT WID	.23	**.57***	.55	.69
MLATII	.51	**.53***	.13	.59
R COMP	.41	.06	**.76***	.75
MLAT III	-.03	.16	**.71***	.53
GROUP ACH	.57	.27	**.61***	.77
MLAT IV	.47	.26	**.51***	.55

Note. Boldface indicates the factor on which the test or subtest had the highest loading.

MLAT I = Modern Language Aptitude Test - Number Learning subtest;
MLAT V = Modern Language Aptitude Test - Paired Associates subtest;
PPVT-R = Peabody Picture Vocabulary Test-Revised;
WRMT WATT = Woodcock Reading Mastery Test-Revised (Word Attack subtest);
PHON AW = Phoneme Awareness;
WRAT-R SPELL = Wide Range Achievement Test-Revised (Spelling subtest);
WRMT WID = Woodcock Reading Mastery Test-Revised (Word Identification subtest);
MLAT II = Modern Language Aptitude Test - Phonetic Coding subtest;
R COMP = Reading Comprehension;
MLAT III = Modern Language Aptitude Test -Spelling Clues subtest;
GROUP ACH = Group Achievement; and
MLAT IV = Modern Language Aptitude Test - Words in Sentences subtest.

Phonological Coding/Recoding dimension, and Factor 3 (20.16%) as a Cognitive Speed Plus dimension.

Factor 1 obtained substantial loadings (>.50) from 3 of the 12 tests and was labeled as defining the dimension Verbal Memory. This factor appeared to measure the cognitive dimension related to verbal memory and vocabulary for both native language and the FL and received heavy loadings from measures of verbal rote memory, oral language comprehension, and vocabulary involved in English and FL courses (Carroll, 1993).

Factor 2 obtained substantial loadings from 5 of the 12 tests and appeared to define the dimension Phonological Coding/Recoding. This factor appeared to measure the awareness that spoken words are composed of sounds, the ability to segment and manipulate sounds within words (both native and foreign), and the ability to decode written words (real words in the native language, FL words, nonsense words in the native language). This dimension addressed both the "meta" aspects of language at the sound level and skill in reading low frequency, unfamiliar, and difficult words to which the student has received little or no previous exposure. This dimension received heavy loadings from measures of phonemic awareness, FL phonology/orthography, and native language phonology/orthography.

Factor 3 obtained substantial loadings on 4 of the 12 tests and appeared to define the dimension Cognitive Speed Plus. This factor appeared to be related to both speed and language processing and production. Cognitive Speed may be described as "quickness in identifying elements, or distinguishing between elements, of a (visual) stimuli pattern, particularly when measured under pressure to maintain focused attention" (Horn, 1988, p. 666). Another aspect of Cognitive Speed involved the "quickness in deciding on answers"—a measure of "just how quickly one produces answers, both correct and incorrect, to problems of moderate difficulty" (Carroll, 1993, p. 615). In a review of 60 years of factor analytic studies, Carroll (1993) included the language processing dimensions of semantic processing speed (i.e., speed of verbal reception), word fluency (i.e., speed of word retrieval), verbal ability (i.e., general verbal knowledge), and semantic fluency (i.e., the speed of idea reception and production). Factor 3 was reflective of the language processing aspects of Cognitive Speed Plus and received heavy loadings from timed measures of FL syntax and native language phonological/orthographic processing, general vocabulary, and general knowledge (e.g., information in reading comprehension paragraphs).

A tolerance diagnostic was computed for each dependent variable and principal component to check for multicollinearity. The tolerance

values, which ranged from .30 to .70, indicated that multicollinearity was not a threat in this data set (Tabachnick & Fidell, 1996).

Four multiple-regression analyses, which examined the relationship among the three Factor Scores and the four outcome variables, were performed with the following results.

FL Grades

In the prediction of first year FL grades, the multiple regression analysis yielded a solution that resulted in an R^2 of .381 and an adjusted R^2 of .361, $F(3, 92) = 18.87$, $p = .0001$. The regression indicated that all three Factor Scores, Verbal Memory, $F(1, 92) = 5.31$, $p = .0001$, Phonological Coding/Recoding, $F(1, 92) = 2.99$, $p = .004$, and Cognitive Speed Plus, $F(1, 92) = 4.41$, $p = .0001$, were significant in predicting first year FL grades (see Table 2).

English Grades

In the prediction of eighth grade English grades, the multiple regression analysis provided a solution that resulted in an R^2 of .289 and an adjusted R^2 of .266, $F(3, 92) = 12.45$, $p = .0001$. The regression indicated that all three Factor Scores, Verbal Memory, $F(1, 92) = 3.53$, $p = .0005$, Phonological Coding/Recoding, $F(1, 92) = 3.92$, $p = .0002$, and Cognitive Speed Plus, $F(1, 92) = 3.03$, $p = .0032$, were significant in predicting eighth grade English grades (see Table 3).

FL Word Recognition

In the prediction of FL word recognition, the multiple regression analysis yielded a solution that resulted in an R^2 of .353 and an adjusted R^2 of .331, $F(3, 92) = 16.67$, $p = .0001$. Two of the three regression coefficients, Verbal Memory, $F(1, 92) = 4.77$, $p = .0001$, and Phonological Coding/Recoding, $F(1, 92) = 5.11$, $p = .0001$, were significant in the prediction of FL word recognition proficiency (see Table 4).

Overall FL Proficiency

In the prediction of overall FL proficiency (FL Total Test), the multiple regression analysis yielded a solution that resulted in an R^2 of .201 and an adjusted R^2 of .175, $F(3, 92) = 7.73$, $\underline{p} = .0001$. Two of the

Table 2
Regression Model Using Factor Scores from Factor Analysis in the Prediction of First Year FL Grades

Factor	Standard Error	*B*	*t* value	*p*
Verbal Memory (Factor 1)	.06	.44	5.31	.0001
Phonological Coding/Recoding (Factor 2)	.06	.25	2.99	.0035
Cognitive Speed Plus (Factor 3)	.06	.36	4.41	.0001

Table 3
Regression Model Using Factor Scores from Factor Analysis in the Prediction of Eighth Grade English Grades

Factor	Standard Error	*B*	*t* value	*p*
Verbal Memory (Factor 1)	.053	.31	3.53	.0005
Phonological Coding/Recoding (Factor 2)	.053	.34	3.92	.0002
Cognitive Speed Plus (Factor 3)	.053	.27	3.03	.0032

Table 4
Regression Model Using Factor Scores from Factor Analysis to Predict FL Word Recognition

Factor	Standard Error	*B*	*t* value	*p*
Verbal Memory (Factor 1)	.30	.39	4.77	.0001
Phonological Coding/Recoding (Factor 2)	.30	.43	5.11	.0001
Cognitive Speed Plus (Factor 3)	.30	.09	1.11	.269

Table 5
Regression Model Using Factor Scores from Factor Analysis to Predict Overall FL Proficiency

Factor	Standard Error	*B*	*t* value	*p*
Verbal Memory (Factor 1)	.72	.37	4.02	.0001
Phonological Coding/Recoding (Factor 2)	.72	.03	.38	.706
Cognitive Speed Plus (Factor 3)	.72	.25	2.63	.01

three regression coefficients, Verbal Memory, $F(1, 92) = 4.02$, $p = .0001$, and Cognitive Speed Plus, $F(1, 92) = 2.63$, $p = .01$, were significant in the prediction of overall FL proficiency after two years of FL instruction (see Table 5).

Discussion

Results of the factor analysis on the combined data sets yielded three components. Together, the three factors accounted for a substantial percentage of the variance (63.28%) in FL proficiency. One of the factors, Verbal Memory (Factor 1), was represented by measures of oral language comprehension, vocabulary, and verbal rote memory. The authors speculated that Factor 1 was representative of the semantic (meaning) and verbal memory aspects of language. Both John Carroll's (1962) model of FL aptitude (i.e., phonetic coding, grammatical sensitivity, inductive language learning ability, rote memory) and Sparks and Ganschow's Linguistic Coding Differences Hypothesis (i.e., phonology/orthography, syntax, semantics) (Sparks, 1995; Sparks & Ganschow, 1991, 1993a,b, 1995a; Sparks, Ganschow, & Pohlman, 1989) emphasize that students' ability to use and understand language, generally, is likely to be the primary determinant of their ability to become proficient in the oral and written aspects of a FL. Whereas both Carroll's and Sparks and Ganschow's models of FL learning explicitly include semantics as necessary components of language proficiency, Carroll's model also includes rote memory.

All three factors appeared to measure conceptually different aspects of language proficiency, given their emergence as separate components in the factor analysis. Phonological Coding/Recoding (Factor 2) was represented by measures of phonemic awareness (the Phoneme Deletion and Pig Latin tasks), FL phonology/orthography (MLAT Phonetic Script subtest), and native language phonology/orthography (word recognition, pseudoword reading, spelling). Cognitive Speed Plus (Factor 3) was represented by *timed* measures of academic achievement (standardized group achievement tests), reading comprehension (standardized reading comprehension tests), and syntax (MLAT Words in Sentences subtest), and phonological/orthographic processing/vocabulary (MLAT Spelling Clues subtest). Although both Factors 2 and 3 were represented by measures of native language and FL skills, the authors speculated that they were separate components in the factor analysis because the testing measures in Factor 2, Phonological Coding/Recoding, all involved the coding and recoding of phonological/orthographic information, whereas the testing measures in Factor 3, Cog-

nitive Speed Plus, all included a speed dimension.

Two of the measures in Factor 2, word recognition and spelling, are phonological coding tasks. Two of the other measures in Factor 2, pseudoword reading and MLAT Phonetic Script subtest, were hypothesized by the authors to be "phonological recoding" tasks. Phonological recoding is defined as "translating letters into sounds by application of letter-sound rules and then recognizing the identities of words from their pronunciations" (Ehri, 1992, p. 107). Gough (1984) suggests that phonological recoding is not used to read words to which the reader has had frequent exposure (e.g., word recognition in a student's native language). Instead, phonological recoding is used primarily to read nonsense words, low frequency words, unfamiliar words, and difficult words that have not received sufficient exposure to enter memory. Pseudoword (nonsense word) reading, then, was determined to be a phonological recoding task.

The MLAT Phonetic Script subtest was also hypothesized to meet the criteria for phonological recoding. On this task, students are asked to listen to the examiner (on a prerecorded tape) read aloud four words, each of which has a pronunciation that is different from the English sound/symbol system (e.g., *tik, tiyk, tis, tiys* are pronounced as /tik/, /tek/, /tis/, /tes/). After reading aloud five sets of four words each, the examiner returns to the first set and instructs the student to choose the one word (out of four) that is subsequently said aloud in each set. The Phonetic Script subtest includes six sets of five words each and introduces new sounds in each subsequent set. In a previous study that used a factor analysis procedure, the MLAT Phonetic Script subtest loaded on a separate component with other MLAT subtests (Sparks, Ganschow, & Patton, 1995). In another study a phonemic awareness measure, the Lindamood Auditory Conceptualization Test, loaded on a component that included the MLAT Phonetic Script subtest and three other MLAT subtests (Ganschow, et al., 1992). In the two aforementioned studies, pseudoword reading, as measured by the WRMT-R: Word Attack subtest, loaded on a native language phonology/orthography component. In the present study, phonological recoding tasks combined with a metaphonological task (phonemic awareness, i.e., Phoneme Deletion and Pig Latin) and emerged as a component important for FL learning.

In contrast to the Verbal Memory and Phonological Coding/ Recoding factors, two of the tasks in Cognitive Speed Plus (Factor 3), group achievement and reading comprehension, appeared to rely directly on students' processing and production of their native language under time constraints. The third testing measure in Cognitive Speed

Plus, the MLAT Spelling Clues subtest, relies directly on students' knowledge of English vocabulary and indirectly on the sound and sound-symbol system of English; this task is also administered under time constraints. On the Spelling Clues subtest, the student is presented with misspelled shortened versions of English words (e.g., luv, ernst, mblm, sidr). The instructions tell the student that the words are not spelled in the usual way, but are spelled approximately as they are pronounced; however, the words adhere to the English sound/symbol system although the students are not told this directly. The student must decode each word before finding a word from a list of five that corresponds most closely in meaning to the target word (e.g., student decodes *luv*, then chooses the word closest in meaning to *love* from the following words: *carry, exist, affection, wash, spy*). Previous studies suggest that success on the MLAT Spelling Clues subtest was likely to be equally dependent on students' knowledge of phonology/orthography as on their vocabulary knowledge (Ganschow, et al., 1992; Sparks, Ganschow, & Patton, 1995). The fourth measure in the Cognitive Speed Plus factor, MLAT Words in Sentences subtest, assessed grammatical sensitivity as well as the ability to process language under time constraints. Although not directly related to Carroll's model of FL learning, the factor Cognitive Speed Plus is similar to the "broad cognitive speed" factor in Carroll's (1993) three-stratum model of cognitive abilities. Findings show that the Cognitive Speed Plus factor appears to be related to those language tasks such as word retrieval and verbal fluency that are performed under time constraints. Moreover, Cognitive Speed Plus includes word retrieval and verbal fluency skills.

Results of the multiple regression analysis showed that all three factors were significant in predicting eighth grade English grade and end-of-year FL grades. This finding indicated that there were no differences in the factors predicting native language and FL classroom performance. The findings suggest that achievement in school-based FL language instruction relies, at least in part, upon the skills measured by the three factors.

In the prediction of FL word recognition, the results indicated that the Verbal Memory and Phonological Coding/Recoding factors were predictive of the ability to "crack the code" in order to read both native and FL words. One would have likely predicted that the several testing measures in Phonological Coding/Recoding (Factor 2), pseudoword reading, native language word recognition, spelling, phonemic awareness, and phonological/orthographic learning in a FL (i.e., MLAT Phonetic Script subtest), would be predictive of students' ability to read words in a FL. However, the authors had not anticipated that the tasks

in Verbal Memory (Factor 1), verbal memory (MLAT Number Learning and MLAT Paired Associates) and receptive vocabulary (PPVT-R), would be predictive of the ability to read words in a FL.

One possible explanation for this finding is that students with stronger phonological/orthographic skills are those who read more frequently. Numerous researchers have found that exposure to print (i.e., reading) can predict a variety of behavioral outcomes even when stringent controls for background characteristics and general cognitive ability are utilized (see Anderson, Wilson, & Fielding, 1988; Guthrie, Schaefer, & Hutchinson, 1991; Stanovich & Cunningham, 1992). One positive behavioral outcome of more frequent exposure to print can be a stronger vocabulary (e.g., see Cunningham & Stanovich, 1991; Stanovich, 1993; West, Stanovich, & Mitchell, 1993). Thus, it is likely that students with strong phonological/orthographic skills (i.e., pseudoword reading, native language word recognition, spelling, phonemic awareness, FL word recognition) read larger amounts of text and subsequently acquire larger vocabularies as a result of reading. In the present study, then, recognition of FL words also might have been affected by the extent of students' exposure to print in their native language.

Results of the multiple regression analysis showed that two of the three factors, Verbal Memory (Factor 1) and Cognitive Speed Plus (Factor 3), were predictive of overall oral and written FL proficiency; Phonological Coding/Recoding (Factor 2) was not significant in the model.[7] We speculate that Phonological Coding/Recoding was not significant in predicting overall FL proficiency because native language word recognition and pseudoword reading as well as metaphonological skill (i.e., phonemic awareness) rely heavily on one's skills in phonological/orthographic processing. Research in native language reading has shown that word recognition and spelling skills are dependent on phonemic awareness and phonological/orthographic processing skills and that variation in phonological processing skill is the primary mechanism that determines early success in reading (see Brady & Shankweiler, 1991; Stanovich, 1992). These skills are acquired early in learning to read and spell (i.e., in kindergarten and first grade), especially for stronger readers who are more likely to be enrolled in FL courses in secondary school. Findings suggest that although reading and spelling FL words is important for FL learning, overall proficiency in a FL, at least for secondary level students (i.e., tenth and eleventh graders), may be more heavily dependent upon the skills measured by the Verbal Memory and Cognitive Speed Plus (i.e., language production and comprehension, vocabulary, verbal rote memory, speed of

language processing, or language skills that can be measured by paper-pencil tasks).

An important finding of this study, is that Factor 2, Phonological Coding/Recoding, was found to be predictive of end of first-year FL *grades*, but not predictive of FL *proficiency* after two years of study. FL educators have speculated that FL grades do not necessarily reflect how well students can read, write, speak, and listen to the FL (i.e., proficiency). There are several reasons why their speculation may be accurate. One reason that FL grades but not FL proficiency would be predicted by phonological/orthographic skill is that educators sample more reading and writing in daily FL classroom work (e.g., reading the textbook, taking written tests and quizzes, spelling words) than is sampled on a FL proficiency test (i.e., listening and speaking in addition to reading and writing).

A second reason Phonological Coding/Recoding may be predictive of FL grades but not FL proficiency is that FL proficiency measures are more holistic in nature. That is, FL proficiency measures are developed and scored in such a way that they do not measure the discrete skills that FL teachers generally sample when assigning grades.

A third reason why the Phonological Coding/Recoding factor might be predictive of FL grades but not FL proficiency may be that the large majority of the study's participants generally had well-developed phonological/orthographic skills in both their native language and the FL. In another study involving the same 96 students (Sparks, Ganschow, et al., 1998), findings showed that their mean scores on the phonological/orthographic measures used in this study were in the higher end of the average range. Ehri (1985) hypothesizes that *written* language development positively enhances *oral* language development because written language serves as a "visual-spatial model for speech" and that acquisition of a written language system "works various changes on spoken language, particularly at the phonetic and lexical levels" (p.361). Thus, the secondary level students with well-developed phonological and orthographic skills may have been able to rely on their skills in, e.g., language processing and production, verbal rote memory, and vocabulary, because they did not have to struggle to read or spell new and unfamiliar FL words in the FL classroom. However, other students with significantly lower levels of phonological/orthographic processing skill may not be able to process and produce language in the FL or learn FL vocabulary words as easily as students with stronger phonological/orthographic processing skills because students with lower levels of phonological and orthographic processing skill have more difficulty with the phonological (i.e., phonemic awareness) and phonological and

orthographic (e.g., word recognition, spelling) aspects of both their native language and the FL. This speculation is consistent with Sparks and Ganschow's Linguistic Coding Differences Hypothesis (Sparks, et al., 1995; Sparks & Ganschow, 1991, 1993b, 1995a). Also, the speculation is intuitively appealing because Sparks, Ganschow, and their colleagues have found that students with significantly lower scores on measures of phonology/orthography (e.g., pseudoword reading, spelling) achieve lower scores on FL proficiency measures than do students with higher scores on phonological/orthographic processing measures (see Ganschow, et al., 1997; Sparks, Artzer, et al., 1998; Sparks, Ganschow, et al., 1998).

Overall, the results of this factor analysis showed that all components of language (phonology/orthography, syntax, semantics, and verbal rote memory) are important for oral and written proficiency in a FL. New to the FL literature is the finding that speed of language processing (i.e., Cognitive Speed Plus) contributed to the variance not only in FL grades, but also in FL proficiency.

Implications

There are several implications of this study. First, students learning to read, write, speak, and listen to a FL are likely to rely on all components of language (i.e., phonological/ orthographic, syntactic, semantic, and verbal memory) to learn the new language. Heretofore, students' ability to use the phonological/orthographic component of language to read (and spell) new and unfamiliar words (i.e., the FL) has not always been considered as important as the other language components in FL instruction. Its importance is reflected in the finding that this component appeared to measure a conceptually different aspect of language proficiency and was a significant predictor of end-of-year FL grades. Students may benefit from FL instruction that emphasizes not only the meaning aspects of language, but also the new sound-symbol system of the FL (see Ganschow & Sparks, 1995; Sparks, Artzer, et al., 1998; Sparks, Ganschow, Artzer, & Patton, 1997).

Second, FL proficiency may rely not only on previous language learning (i.e., the components of one's native language), but also on the ability to learn unfamiliar components of a new language quickly and effectively, i.e., speed of language processing and production. In this study, speed proved to be important to students' FL grades and to their oral and written proficiency in a FL.

Third, the lack of a separable FL aptitude component (e.g., a component composed solely or primarily of MLAT subtests) in the

present study suggests that FL aptitude (i.e., potential to learn a FL) may be an extension of native language skills. This finding is consistent with and supportive of the Linguistic Coding Differences Hypothesis (Sparks, 1995; Sparks & Ganschow, 1991, 1993b, 1995a; Sparks, Ganschow, & Pohlman, 1989), which is conceptually similar to Carroll's (1962) model of FL aptitude embodied in the MLAT, and to Carroll's (1993) model in which he finds that FL aptitude (which was associated with phonetic coding, grammatical sensitivity, spelling, and verbal language) was subsumed by a "crystallized" intelligence factor. The finding is also supportive of Carroll's (1973) speculation that FL aptitude is a "residue" of native language skills.

Fourth, the findings suggest that emphasizing the written (i.e., reading, writing) aspects of a new language along with the oral (i.e., listening, speaking) aspects may be important for effective FL learning, especially with students who have histories of and/or current difficulty with recoding and spelling in their native language (i.e., phonology/orthography). Findings are supportive of Ehri's (1985) hypothesis that written language development positively enhances the development of oral language skills.

Finally, FL educators may want to further examine the relationship between grades assigned in classroom FL courses and oral and written proficiency in a FL. Recent research by Sparks, Ganschow, et al. (1998) has shown that students who achieve higher end-of-year FL grades (i.e., As and Bs) in first and second year FL courses have significantly stronger overall proficiency in a FL than students who achieve lower end-of-year FL grades (i.e., Cs, Ds, and Fs).

Some additional research questions that might be investigated are: Do grades in FL courses reflect skill in reading and writing the FL more than speaking/listening to the FL? Do grades in FL courses correlate more strongly with written or oral FL proficiency?

Notes

[1]The authors have previously changed the name of their hypothesis from "deficit" to "differences" to reflect the notion that FL learning skill occurs along a continuum from very good to very poor FL learners.

[2]In other investigations, the same participant sample was used (Sparks, et al., 1997, 1998). However, different analytic techniques, i.e., group comparisons and prediction of FL proficiency, not factor analysis, were employed in those studies.

[3]We thank the following individuals for their patience and cooperation in the completion of this study: Joanne Brewer, Carol Dettenwanger, Marilyn Herring, Kim Icsman, Ann Johnson, Nancy Lamb, Ed Rudder, Mary Sies, Shirley Speaks, and Craig Ullery.

. [4]Coefficient Alpha can be applied to both conventional tests where answers are marked "right" or "wrong" and to responses receiving different weights (McDaniel, 1994).

[5]We thank the following individuals for their assistance in administering and scoring the native language and FL aptitude tests: Sue Aielli, Loreli Albus, Diane Beck, Sue Jarvis, Jane Pohlman, Mikki Springer, Kim Stevens, Mary Thompson, and Connie Yoho.

[6]In the varimax rotation, there were a total of two negatively-loaded variables out of 36 variables. Both of these negative loadings were in the -.03 range and did not load significantly on the factors that were extracted.

[7]Although the adjusted R^2 resulted in a modest value of .175, this value does explain 17.5% of the variance in the model and provides useful information in the development of a theoretical model (Krathwohl, 1993).

References

American Council of the Teaching of Foreign Languages. (1986). *American Council of the Teaching of Foreign Language Proficiency Guidelines*. Hastings-on-Hudson, NY: Author.

American Council of the Teaching of Foreign Languages. (1989). *American Council of the Teaching of Foreign Language Proficiency Guidelines*. Hastings-on-Hudson, NY: Author.

Anderson, R., Wilson, P., & Fielding, L. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23,* 285-303.

Au, S. (1988). A critical appraisal of Gardner's social-psychological theory of second language (L2) learning. *Language Learning, 38,* 75-100.

Brady, S., & Shankweiler, D. (Eds.). (1991). *Phonological processes in literacy*. Hillsdale, NJ: Erlbaum.

Brown, J., Bennett, J., & Hanna, G. (1981). *Nelson-Denny Reading Test*. Chicago, IL: Riverside.

Carroll, J. (1962). The prediction of success in intensive foreign

language training. In T. Glaser (Ed.), *Training and research in education* (pp. 87-136). Pittsburgh, PA: University of Pittsburgh Press.

Carroll, J. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics, 2*, 5-14.

Carroll, J. (1993). *Human cognitive abilities: A survey of factor analysis studies*. NY: Cambridge University Press.

Carroll, J., & Sapon, S. (1959). *Modern Language Aptitude Test (MLAT): Manual.* San Antonio, TX: Psychological Corporation.

Catts, H. (1986). Speech production/phonological deficits in reading of disordered children. *Journal of Learning Disabilities, 19*, 504-508.

Crain, S. (1989). Why poor readers misunderstand spoken sentences. In.D. Shankweiler & I. Liberman (Eds.), *Phonology and reading disability* (pp. 133-165). Ann Arbor, MI: University of Michigan Press.

Cunningham, A., & Stanovich, K. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*, 264-274.

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance.

Ehri, L. (1985). Effects of printed language acquisition on speech. In D. Olson, N. Torrence, & A. Hildyard (Eds.), *Literacy, language and learning* (pp. 333-367). Cambridge, MA: Cambridge University Press.

Ehri, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale, NJ: Erlbaum.

Ehrman, M. (1990). The role of personality type in adult language learning: An ongoing investigation. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 126-178). Englewood Cliffs, NJ: Prentice-Hall.

Ehrman, M., & Oxford, R. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 76*, 67-89.

Ganschow, L., & Sparks, R. (1991). A screening instrument for foreign language learning problems: Evidence for a relationship between native and second language learning. *Foreign Language An-*

nals, 24, 383-398.

Ganschow, L., & Sparks, R. (1995). Effects of direct instruction in Spanish phonology on the native language skills and foreign language aptitude of at-risk foreign language learners. *Journal of Learning Disabilities, 28,* 107-120.

Ganschow, L., & Sparks, R. (1996). Foreign language anxiety among high school women. *Modern Language Journal, 80,* 199-212.

Ganschow, L., Sparks, R., Anderson, R., Javorsky, J., Skinner, S., & Patton, J. (1994). Differences in anxiety and language performance among high- and low-anxious college foreign language learners. *Modern Language Journal, 78,* 41-55.

Ganschow, L., Sparks, R., Javorsky, J., Pohlman, J., & Bishop-Marbury, A. (1991). Identifying native language difficulties among foreign language learners in college: A "foreign" language learning disability? *Journal of Learning Disabilities, 24,* 530-541.

Ganschow, L., Sparks, R., Patton, J., & Javorsky, J. (1992). Factors relating to learning a foreign language among high and low risk high school students and students with learning disabilities. *Applied Language Learning, 3,* 37-63.

Gardner, R. (1985). *Social psychology and second language learning*. London: Arnold.

Gardner, R. (1990). Attitudes, motivation, and personality as predictors of success in foreign language learning. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 179-221). Englewood Cliffs, NJ: Prentice-Hall.

Gardner, R., & Lambert, W. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Gough, P. (1984). Word recognition. In P. Pearson, R. Barr, M. Kamil, & P. Rosenthal (Eds.). *Handbook of reading research* (pp. 225-253). New York: Longman.

Guthrie, J., Schaefer, W., & Hutchinson, S. (1991). Relations of document literacy and prose literacy to occupational and societal characteristics of young black and white adults. *Reading Research Quarterly, 26,* 30-48.

Hall, J., & Humphreys, M. (1982). Research on specific learning disabilities: Deficits and remediation. *Topics in Learning and Learning Disabilities, 2,* 68-78.

High School Placement Test. (1991). Bensenville, IL: Scholastic Testing Service.

Horn, J. (1988). Thinking about human abilities. In J. Nesselreade & R. Cattel (Eds.), *Handbook of multivariate experimental psychology* (pp. 645-685). New York: Plenum.

Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70,* 125-132.

Iowa Tests of Basic Skills. (1990). Chicago, IL: Riverside.

Jastak, S., & Wilkinson, G. (1984). *Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates.

Javorsky, J., Sparks, R., & Ganschow, L. (1992). Perceptions of college students with and without specific learning disabilities about foreign language courses. *Learning Disabilities: Research and Practice, 7*, 31-44.

Koda, K. (1992). The effects of lower-level processing skills on foreign language reading performance: Implications for instruction. *Modern Language Journal, 76,* 502-512.

Krathwohl, D. (1993). *Methods of educational and social science research: An integrated approach*. White Plains, NY: Longman.

Lett, J., & O'Mara, F. (1990). Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defense Language Institute Foreign Language Center. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 222-260). Englewood Cliffs, NJ: Prentice-Hall.

MacIntyre, P., & Gardner, R. (1991). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning, 41,* 85-117.

MacIntyre, P., & Gardner, R. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*, 283-305.

Mann, V., Shankweiler, P., & Smith, S. (1984). The association between comprehension of spontaneous sentences and early reading ability: The role of phonemic representation. *Journal of Child Language, 11,* 627-643.

McDaniel, E. (1994). *Understanding education measurement.* Madison, WI: WCB Brown & Benchmark.

Oller, J. (1981). Research on the measurement of affective variables: Some remaining questions. In R. Andersen (Ed.), *New dimensions in second language acquisition research* (pp. 14-32). Rowley, MA: Newbury House.

Omaggio, A. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston, MA: Heinle.

Oxford, R. (1990a). Styles, strategies, and aptitude: Connections for language learning. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 67-125). Englewood Cliffs,

NJ: Prentice-Hall.

Oxford, R. (1990b). *Language learning strategies: What every teacher should know*. Boston, MA: Heinle.

Parry, T. & Standsfield, C. (Eds.) (1990). *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice-Hall.

Petersen, C., & Al-Haik, A. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement, 36*, 369-380.

Pimsleur, P. (1966). *Language Aptitude Battery*. New York: Harcourt, Brace.

Skehan, P. (1986). The role of foreign language aptitude in a model of school learning. *Language Testing, 3*, 188-221.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition, 13*, 275-298.

Sparks, R. (1995). Examining the linguistic coding deficit hypothesis to explain individual differences in foreign language learning. *Annals of Dyslexia, 45*, 187-214.

Sparks, R., Artzer, M., Patton, J., Ganschow, L., Miller, K., Hordubay, D., & Walsh, G. (1998). Benefits of multisensory language instruction for at-risk learners: A comparison study of high school Spanish students. *Annals of Dyslexia*, 48, 239-270.

Sparks, R., & Ganschow, L. (1991). Foreign language learning difficulties: Affective or native language aptitude differences? *Modern Language Journal,75*, 3-16.

Sparks, R., & Ganschow, L. (1993a). The impact of native language learning problems on foreign language learning: Case study illustrations of the linguistic coding deficit hypothesis. *Modern Language Journal, 77,* 58-74.

Sparks, R., & Ganschow, L. (1993b). Searching for the cognitive locus of foreign language learning problems: Linking first and second language learning. *Modern Language Journal, 77*, 289-302.

Sparks, R., & Ganschow, L. (1993c). The effects of a multisensory structured language approach on the native language and foreign language aptitude skills of at-risk learners: A follow-up and replication study. *Annals of Dyslexia, 43*, 194-216.

Sparks, R., & Ganschow, L. (1995a). A strong inference approach to causal factors in foreign language learning: A response to MacIntyre. *Modern Language Journal, 79*, 235-244.

Sparks, R., & Ganschow, L. (1995b). Parent perceptions in the screening for performance in foreign language courses. *Foreign Language Annals, 28*, 371-391.

Sparks, R., & Ganschow, L. (1996). Teachers' perceptions of students'

foreign language academic skills and affective characteristics. *Journal of Educational Research, 89,* 172-185.

Sparks, R., Ganschow, L., Artzer, M., & Patton, J. (1997). Foreign language proficiency of at-risk and not-at-risk learners over two years of foreign language instruction. *Journal of Learning Disabilities, 30,* 92-98.

Sparks, R., Ganschow, L., Artzer, M., Siebenhar, D., Plageman, M., & Patton, J. (1998). Differences in native language skills, foreign language aptitude, and foreign language grades among high, average, and low proficiency foreign language learners: Two studies. *Language Testing, 15,* 181-216.

Sparks, R., Ganschow, L., Fluharty, K., & Little, S. (1996). An exploratory study on the effects of Latin on the native language skills and foreign language aptitude of students with and without learning disabilities. *Classical Journal, 91,* 165-184.

Sparks, R., Ganschow, L., & Javorsky, J. (1993). Perceptions of high and low-risk students and students with learning disabilities about high school foreign language courses. *Foreign Language Annals, 26,* 491-510.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992a). Test comparisons among students identified as high-risk, low-risk, and learning disabled in high school foreign language courses. *Modern Language Journal, 76,* 142-159.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992b). Identifying native language deficits in high- and low-risk foreign language learners in high school. *Foreign Language Annals, 25,* 403-418.

Sparks, R., Ganschow, L., & Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology, 87,* 638-655.

Sparks, R., Ganschow, L., Patton, J., Artzer, M., Siebenhar, D., & Plageman, M. (1997). Prediction of foreign language proficiency. *Journal of Educational Psychology, 89,* 549-561.

Sparks, R., Ganschow, L., & Pohlman, J. (1989). Linguistic coding deficits in foreign language learners. *Annals of Dyslexia, 39,* 179-195.

Sparks, R., Ganschow, L., Pohlman, J., Artzer, M., & Skinner, S. (1992). The effects of a multisensory, structured language approach on the native and foreign language aptitude skills of high risk, foreign language learners. *Annals of Dyslexia, 42,* 25-53.

Spinelli, E. (1988). *Situation cards for oral evaluation: First year*

French, German, and Spanish. New York: Holt, Rinehart, & Winston.

Stanovich, K. (1988). The right and wrong places to look for the cognitive locus of reading disability. *Annals of Dyslexia, 38,* 154-177.

Stanovich, K. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ: Erlbaum.

Stanovich, K. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. In H. Reese (Ed.), *Advances in child development and behavior, 24*, 133-180. San Diego, CA: Academic Press.

Stanovich, K., & Cunningham, A. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory and Cognition, 20*, 51-68.

Tabachnick, B., & Fidell, L. (1996). *Using multivariate statistics*. New York: Harper Collins.

Tiedemann, J. (1989). Measures of cognitive styles: A critical review. *Educational Psychologist, 24,* 261-275.

Vellutino, F., & Scanlon, D. (1986). Linguistic coding deficits and metalinguistic awareness: Their relationship to verbal and code acquisition in poor and normal readers. In D. Yaden & S. Templeton (Eds.), *Metalinguistic awareness and beginning literacy* (pp. 115-141). Portsmouth, NH: Heinman.

West, R., Stanovich, K., & Mitchell, H. (1993). Reading in the real world and its correlates. *Reading Research Quarterly, 28*, 35-50.

Wilson, K. (1989). Enhancing the interpretation of a norm-referenced second-language test through criterion referencing: A research assessment of experience in the TOEIC testing context. Princeton, NJ: Educational Testing Service.

Wilson, K., & Graves, K. (1995). An assessment of validity-related properties of the secondary level English proficiency test and measures of ESL speaking and writing abilities in the Temple University-Japan context. Unpublished manuscript.

Woodcock, R. (1987). *Woodcock Reading Mastery Test-Revised*. Circle Pines, MN: American Guidance.

Appendix A

Alphabetical List and Descriptions of Testing Instruments

High School Placement Test (HSPT TOT):

This group-administered, comprehensive placement test is designed specifically to aid in the selection and/or placement of students entering high school. The test measures Cognitive Skills (Verbal, Quantitative) and Basic Skills (Reading, Math, Language).

Iowa Tests of Basic Skills, Form J, Level 14 (ITBS TOT):

This test is a standardized measure of comprehensive growth in fundamental academic skills. It consists of subtests of language, reading, vocabulary, and mathematics, with questions presented in a multiple-choice format.

Iowa Tests of Basic Skills, Form H, Level 14 - Reading Comprehesion (ITBS RCOMP):

This test is comprised of paragraphs of varying lengths which the student reads and then answers questions presented in a multiple-choice format.

Modern Language Aptitude Test (MLAT):

This test measures foreign language aptitude using a simulated format to provide an indication of probable degree of success in learning a foreign language; it includes five subtests. The Long Form (MLAT LF) includes all five subtests. The subtests are: *MLAT I: (Number Learning)*: student learns numbers of a made-up language, and then transcribes spoken number words into written digits on hearing them presented rapidly; *MLAT II: (Phonetic Script)*: student listens to a sequence of syllables (many with no meaning in English) while looking at their graphemic transcriptions and is asked to quickly learn how the sounds (phonemes) correspond to the letters (graphemes); *MLAT III: (Spelling Clues)*: student reads English words presented as abbreviated spelling (e.g., luv) and then chooses the one word (out of five) that corresponds most nearly in meaning (e.g., carry, exist, affection, wash, spy); *MLAT IV: (Words in Sentences)*: student reads a “key” sentence in which

a word is underlined, reads another sentence in which five words and phrases are marked as possible choices, and chooses the word or phrase in the second sentence that has the same grammatical function as the marked word or phrase has in the "key" sentence; and *MLAT V: (Paired Associates)*: student memorizes a list of nonsense words with their assigned English meanings.

Nelson-Denny Reading Test (NELSON), Form E:
This test consists of a series of eight paragraphs that measure the ability to read and answer multiple-choice comprehension questions in a timed format.

Peabody Picture Vocabulary Test-Revised (PPVT-R), Form L:
This test measures receptive vocabulary for Standard American English.

Phoneme Deletion (Phoneme Deletion):
This informal phonemic awareness measure has twenty items that test the ability to delete an initial, final, or medial phoneme and form a spoken word.

Pig Latin (Pig Latin):
This informal phonemic awareness measure has fifteen items that test the ability to delete the initial phoneme from a spoken word, move the phoneme to the end of the word, and then add an /a/ sound to the end of the new word. For example, the student must say "lackba" for *black*. The measure is composed of one, two, and three-syllable words.

Wide Range Achievement Test-Revised (WRAT-R SPELL), Spell ing subtest:
This test measures performance on writing single words from dictation.

Woodcock Reading Mastery Test-Revised (WRMT BSK), Form G:
The *Basic Skills Cluster* tests two aspects of reading: *Word Identification* (WRMT WID) tests ability to read isolated words (e.g., urgent, hysterical, causation, heterogeneous); and *Word Attack* (WRMT WATT) tests ability to read (pseudo) nonsense words (e.g., dee, poe, vunhip, mancingful).

Appendix B

Sample Word Recognition Lists for Spanish, French, and German

Spanish		French	
List 1	List 2	List 1	List 2
anoche	agosto	adorable	difficile
enero	efecto	midi	chimie
isla	ideal	robe	bonne
orilla	oriente	haise	caisse
usted	urbano	difficile	adorable
mesa	linda	croix	soie
señora	compañero	quest	coûte
jefa	junta	soeur	jeune
entrenamiento	historiadora	magnifique	peigner
salon	peatón	aéroport	aéroport
inventado	invierno	soleil	sommell
mural	musa	Joël	coïncidence
la amada	la alemana	longtemps	contravention
agencia	agitar	pluie	fruit
filólogo	préstamo	médicament	médicament
alto	algo	rue	vue
corriente	cerradura	hier	mariage
antena	anterior	décidé	télé
repentinamente	cuidadosamente	nationalité	nationalité
extensión	preocupación	réveillon	baigner

German	
List 1	List 2
Zehn	oval
Oktav	Bergkundige
Sieh	Zürich
Jawohl	Baümen
Jodeln	Juni
Deshalb	Tastatur
Postboten	Notiz
Büro	entgegennehmen
Imperfekt	Geschwindigkeit
Deutsch	Begrenzung

Authors

RICHARD L. SPARKS, Associate Professor of Special Education, College of Mt. St. Joseph, Cincinnati, Ohio. *Specializations:* foreign language learning, foreign language learning problems, and reading disabilities.

JAMES JAVORSKY, doctoral candidate in Special Education, Purdue University, West Lafayette, Indiana. *Specializations:* attention deficit disorder, reading disabilities, and foreign language learning problems.

LEANORE GANSCHOW, Professor Emerita of Special Education, Miami University, Oxford, Ohio. *Specializations:* language disabilities (native, foreign, oral, and written).

JON PATTON, Visiting Professor and Senior Research Consultant, Miami University, Oxford, Ohio. *Specializations:* Statistical modeling and consulting about statistical packages.

Applied Language Learning
1998, Vol. 9, Nos. 1& 2, pp. 107-119

Essay Scores as Instruments for Placement and Advancement in an Intensive English Program

Lynne Davis
Ruth Johnson
Floyd Olive
Southern Illinois University

Can essay scores of English-as-a-second-language (ESL) students be used as reliably for placement and/or advancement purposes as their Test of English as a Foreign Language (TOEFL) scores? Can these essay scores be used as reliably for placement and/or advancement purposes as recommendations for placement and/or advancement made by classroom teachers? What is the relationship between Test of Written English (TWE) sample essay scores and TOEFL scores when they are used to make decisions about ESL students' progression from one level to another within an intensive English program (IEP)? And what is the relationship between TWE sample essay scores and teacher recommendations when these are used to make placement and/or advancement decisions?

Exams such as the TOEFL that are usually used for placement, including placement into writing classes, are advantageous because they provide quick evaluation and objectivity (Leki, 1991; Perkins, 1983; Perkins & Pharis, 1977). Those who argue in favor of using such indirect measures of writing ability, such as measures of grammar, usage, word choice and syntax, say such tests actually measure the construct also measured by essay exams (in this case, English-language writing ability); that is, objective tests of writing have construct validity (Educational Testing Service, 1992). They also maintain that an objective, indirect measure of writing ability is superior to an essay test; it prevents the test taker from using avoidance strategy to evade using constructions he or she does not know or is unsure of using (Diederich, 1974).

© 1998, Lynne Davis, Ruth Johnson, and Floyd Olive

Having noted these supporting arguments for such objective tests, it is important also to note that objective tests have been criticized for not testing skills necessary for classroom performance (Leki, 1991). Critics argue in favor of an essay exam, saying that a writing test is the appropriate instrument for measuring writing ability/skill (Kroll, 1991; White, 1993). In addition, they say that essay scores reveal not only writing ability, but also "a robust measure of global language proficiency" (Kaczmarek, 1977, p. 159). This finding, says Leki (1991), is crucial when testing ESL students because it is not only their writing ability that must be determined, but also their general proficiency in English in order to place them appropriately into writing classes.

It should be noted here regarding the holistic grading method that recent discussions question the value of holistic scoring, as used in the TWE, for example, in essay and essay test assessment (Hamp-Lyons, 1995). The disadvantages to essay tests are that, because a time limit is imposed and because the students are writing in their second language, the sample may be very short. There are also concerns with the prompts provided; one concern that affects construct validity is the comparability of the various topic types used on the test (Stansfield & Ross, 1988). Some argue that restricted-time essay tests may produce writing that does not resemble the writing that the participant would produce under process-writing conditions (Caudery, 1990). Further, writing samples are time-consuming to score and the scoring is more subjective. Judging writing is precariously like grading personalities. One rater may consider the positive value of a certain characteristic to be so predominant in an essay as to overshadow a multitude of shortcomings, where another rater may disagree. Vaughan (1991) notes that while holistic assessment clearly works for some essays, in many cases holistic assessment does not work, and in these borderline cases, raters may well follow their own styles of judgment. Essay scoring also requires more expertise on the part of the scorers (Leki, 1991) although Kaczmarek (1977) says that both subjective and objective essay evaluation methods work well and each correlates highly with other valid ESL proficiency measures.

Research Question

Many IEPs use the TOEFL score to make admissions decisions (Leki, 1991) and for placement purposes. It is assumed that if an incoming ESL student achieves a certain score on the TOEFL, then she or he is ready for the content of a certain level in the IEP. And, if an ESL student is continuing in an IEP, then the TOEFL score is factored

with the recommendations of the student's teacher(s) for placement. However, it is generally recognized that the TOEFL is not a valid measure of a student's position within a particular IEP (Leki, 1991). Instead, placement decisions should be matched with assessment instruments related to the IEP program curriculum. For this reason, the Center for English as a Second Language (CESL) sought measures that would be valid for placement decisions within its curriculum. Also, the TOEFL includes no sample of student writing and recommendations from teachers do not consistently include considerations of student writing.

Recently, Educational Testing Service (ETS) has required TOEFL testing sites to administer sample TWE essay tests. The TWE was developed based on survey data which revealed that faculty believed writing has a major role in the academic community and tests of writing should be based on what students are expected to produce (Kroll, 1991). Initially, the primary purpose of these tests is to train scorers so that TWE can become a part of TOEFL. (As of May 1990, ETS reported that the variance accounted for by the TOEFL section scores predicting the TWE score ranged from .80 to .84 (n=91,146) (DeMauro, 1993).

In addition, however, TWE sample essays can be used to make placement decisions. For example, at the time of data collected for this study, at the CESL, decisions regarding placement were made using TOEFL scores and/or teacher recommendations as information sources. The TOEFL was required when a student entered and exited the program. Course grades were used to determine progress through the program. A passing grade of 75% in a writing class, based on an average of in-class and out-of-class writing as well as a timed final exam, would allow a student to pass to the next writing class. (Split-level placement was then used.) Progress in other skills, such as grammar, reading, and speaking, was measured independently with the same score—75%—required for passing. However, using essays like the TWE to help determine placement is emerging as an option. At present, scores on the TWE are not used in placement decisions, and split-level placement has been abolished. The central question in this research study is whether to use writing samples as a means for determining placement and/or advancement in addition to relying on scores from objective tests and/or teacher recommendations.

Method

Participants

The participants were 107 IEP students in CESL levels 1 (beginning) to 4 (advanced) enrolled in Summer 1993; 96 continuing students, originally placed by TOEFL score and advanced by passing course work; and 41 new students, placed by TOEFL.

Scores from 105 participants were used; 2 were dropped because of disagreement in raters' scores after 5 readings by 4 raters.

Test

The instrument used was the English Essay Pretest (EEP) from ETS, which was used as the pilot instrument for gathering reliable prompts for the initial versions of the TWE. So for the purposes of this study the terms EEP and TWE are interchangeable. The test was given in the third term (May-July) of the 1993 academic year. The test consists of a thirty-minute writing sample based on a single writing prompt. The graders used the 6-point criterion-referenced TWE scoring guide.

Raters

The raters were one applied linguistics professor, with 6 years of ESL teaching experience, and two IEP teachers, with 18 and 9 years of ESL teaching experience, respectively. All had had experience with teaching ESL writing and all were trained in the scoring of EEP/TWE essays. A fourth experienced rater was involved in the grading of the essays only during the last round.

Preliminary Calibration Set

The raters initially holistically test graded 7 final essays from CESL students written during the previous year and discussed their scores; differences were resolved based on the criterion-referenced scale. The actual grading was then done.

A grade correspondence as used in TWE ratings was implemented; that is, interrater agreement depended on two readers assigning the same whole number scores or +/- one whole number score to the essays (Kroll, 1991).

Grading of the Essays

For the first reading, the essays were divided equally among the raters, who read them independently, using the TWE scoring guide and the sample essays for reference. A second round of independent readings was done, with scores from the first round being kept secret. Comparisons of the first and second readings yielded these results: Two raters had agreed on the scores for 94 of the essays (88%). The group of essays for which the 3 readers did not agree within +/- 1-point were given to the rater who had not read them (that is, if Raters #1 and #2 had not agreed on the score for essay #1, then that essay was given to Rater #3 to score). Of the 13 essays that required a third reading, all but 2 of them received a score from the third rater that was within +/- 1 point of one of the other two raters. The remaining 2 were sent to a fourth independent reader, but were finally removed from consideration in the results because no clear consensus could be reached regarding their scores.

Interrater Reliability

Because the essays were read in successive readings, an interrater reliability score for each round of readings was calculated. The first and second rounds combined produced an interrater reliability score of .879 and the third round, .846. The average interrater reliability score over the readings was .862.

Essay Score and CESL Level Correspondence

Correspondence between the EEP grading rankings and CESL levels was established by the two raters who are also instructors in the IEP (CESL). Although the instructors may have taught some of the participants and made recommendations for them, the evaluations of the EEP essays, which were known by number only, were rated on the TWE/EEP scale for which the raters established interrater reliability during the calibration session. Having finished grading according to the TWE/EEP scale, the two raters who are the teachers in the IEP reviewed the sample essays and decided that EEP scores 1 through 4 corresponded to placement into CESL levels 1 through 4, respectively; an EEP score of 5 would place a student into undergraduate study or into CESL level 5, for students preparing for graduate study, whichever was appropriate for the student. A student who scored 6 would be proficient for graduate study.

Results

The correlation analysis for overall placement (independent variable: X) and EEP score (dependent variable: Y) is presented in Table 1. The correlation coefficient (r) is .58; thus, the placement of the student into a class accounts for 34% of the variance in the EEP score (r-squared), with a p-value of .0001, significant at the .05-level of alpha.

Table 1
Correlation of Placement (overall) and Writing Score (N=105)

Simple Statistics						
Variable	Mean	Std Dev	Min score	Max score	Corr	p
Overall placement	3.06	1.00	1	4		
Writing score	2.45	0.97	1	5		
Pearson correlation coefficient*					0.58	0.001

*H^0: Rho=0

The correlation analysis for placement by teacher recommendation (independent variable: X) and EEP score (dependent variable: Y) is presented in Table 2. The correlation coefficient (r) is .62; thus, the placement of the student into a class accounts for 38% of the variance in the EEP score (r-squared), with a p-value of .0001, significant at the .05-level of alpha.

Table 2
Correlation of Placement by Teacher Recommendations and Writing Score (N=63)

Variable	Simple Statistics Mean	Std Dev	Min score	Max score	Corr	p
Placement by teacher recommendation	2.60	0.96	1	5		
Writing score	3.40	0.73	1	4		
Pearson correlation coefficient*					0.62	0.001

*H^0: Rho=0

The correlation analysis for TOEFL score (independent variable: X) and EEP score (dependent variable: Y) is presented in Table 3. The correlation coefficient (r) is .55; thus, the placement of the student into a class accounts for 30% of the variance in the EEP score (r-squared), with a p-value of .0002, significant at the .05-level of alpha.

Table 3
Correlation of Placement by TOEFL Score and Writing Score (N=41)

Variable	Simple Statistics Mean	Std Dev	Min score	Max score	Corr	p
Placement by TOEFL score	423.36	63.28	308	543		
Writing score	2.19	0.95	1	1		
Pearson correlation coefficient*					0.55	0.001

*H^0: Rho=0

The correlations for the EEP with overall placement, with placement based on teacher recommendation, and with placement based on TOEFL score are statistically significant and moderately high. What is of additional interest is how the EEP would have fared as a predictor of class level placement. The researchers had predetermined the correspondence between the EEP score and the CESL level (see p. 111).

Discussion

The results of this study suggest that the TWE, or a similar essay exam, is a useful instrument for placement into and advancement within an IEP. In a good percentage of the cases, the EEP score was a more conservative measure than was either the TOEFL score or teacher recommendation when used to place or to advance students (Tables 4-6).

In the majority of TOEFL-score cases (17), students were placed by their essay scores at the same level as their placement by their TOEFL scores. The second largest group (11) placed one level below their placement by their TOEFL score on the basis of their essay scores. Other students were placed at one level higher than placement by TOEFL score based on their essay scores (7), 2 levels lower (6), and 2 levels higher (1).

However, the substantial number of cases in which students placed one level lower than TOEFL placement on the basis of their writing seems to reflect what is known from the experiences of classroom teachers: Producing written English is often a more difficult and complex skill than is choosing the correct form from several provided. The writing skill is often on the same level with overall English-language ability (Kaczmarek, 1977), but almost as often lags behind. This is a fairly common observation made by the researchers in this study.

In the case of placement based on teacher recommendation versus placement by an essay score, an even higher correlation was found, enhancing the reliability of the essay score as a predictor of global language ability. Again, though, placement of students according to essay scores was often at levels lower than placement according to coursework grades and teacher judgment. Placement according to essay test score was at the same level as teacher placement in 20 cases, and one level below teacher placement in 33 cases. Essay scores placed 2 students one level higher than did their teachers' recommendations; 7 students two levels lower than teacher recommendations, and one student three levels lower.

Table 4
Comparison of Placement of Students Using Writing Score Only Versus Actual Placement (N=105)

	Writing Score				
Actual class placement	1	2	3	4	5
1	6	3	0	0	0
2	10	9	2	1	0
3	0	18	10	3	0
4	1	13	16	12	1

Table 5
Comparison of Placement of Students Using Writing Score Only Versus Actual Class Placement by TOEFL Score (N=42)

	Writing Score				
Actual class placement	1	2	3	4	5
1	6	3	0	0	0
2	4	5	2	1	0
3	0	4	4	2	0
4	0	6	3	2	0

Table 6
Comparison of Placement of Students Using Writing Score Only Versus Class Placement by Teacher Recommendation (N=63)

	Writing score				
Actual class placement	1	2	3	4	5
1	0	0	0	0	0
2	6	4	0	0	0
3	0	14	6	1	0
4	1	7	13	10	1

If the phenomenon of a lower proficiency in writing is taken as a normal, consistent, reliable feature, as it appears to be from these results, then the discrepancy in TOEFL placements and placements by essay scores can be expected. It may be that teachers and administrators will recommend this lower placement. Or, only the essays that fall below the standard one level lower than TOEFL placement or above TOEFL placement level will be taken into account. In either case, the essays will be given serious consideration as reliable indicators and concrete evidence of the student's ability to use the language.

The issue of the two essays that were excluded from the study because the three raters could not reach agreement on scoring them deserves special consideration; namely, what was characteristic of these essays that rendered them "good" in the eyes of one rater and "not good at all" in the eyes of another? The researchers took a closer look at these two essays to understand the process of scoring these two essays in particular. The first essay received the following scores from the raters: Rater #1: 3; Rater #2: 2; Rater #3: 4. These scores held for a second round of reading both within a month of the original readings and a year afterwards.

Comments from Rater #3, who gave the essay the highest score, were that the essay was well-organized and content was treated logically; the essay was written with good grammar; and there were "no terrible handwriting or spelling obstacles." Rater #1, who scored in the middle, commented that there was only a weak thesis statement which affected the essay's organization; sentence structure and phrasing were wordy and awkward; one of the main points in the essay did not have

enough detailed support; and word choice was weak in places. Rater #2, who gave the essay the lowest score, said that word choice was simplistic and, at times, incorrect; many run-on sentences appeared; there was no evidence of knowledge of conventional formatting; and none of the points was developed well.

In the case of the other essay, the following scores were given: Rater #1: 2; Rater #2: 1; and Rater #3: 3. The highest score was given by Rater #3, who commented that the essay was thoughtful, fragmented and, well-organized and it contained significant grammar problems. Rater #1 said that the essay contained no clear thesis, had a lack of transitions and a lack of detail. This rater also noted the sentence fragments, errors in tense, and awkward sentence structure. Rater #2 commented that the writer tackled two big issues in an essay of essentially only two paragraphs and, thus, was underdeveloped. In addition, this rater noted serious grammar errors, including run-on sentences, fragments, shifts in verb tenses, and virtually no use of the article system.

From the results of this study, the researchers concur that, because essay exams require a productive use of language which integrates a wide spectrum of linguistic skills, such exams provide a more robust, albeit more complex, measure of English ability than does a test consisting largely of recognition and response to structures already provided, as is the case with the TOEFL. Writing an essay, like carrying on a conversation, requires the student to recall and integrate vocabulary and structure in order to create meaning. Thus, such a sample of language gives more useful information about the individual's skill in using the language than does an objective test of discrete items.

We agree that the rating process in writing assessment needs further research regarding scales and forms of evaluation and, indeed, more attention paid to the shared values of raters regarding the language and content relationship in writing (Connor-Linton, 1995; Mohan & Low, 1995). At the same time, we would argue that essay scores can provide an acceptable measure for student placement when they are used in combination with TOEFL scores (and/or another measure, perhaps an oral test such as the Foreign Service Interview) and that they add a dimension to assessment that is not available with the use of an objective measure such as the TOEFL alone.

References

Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. *ELT Journal, 44*(2), 122-130.

Connor-Linton, J. (1995). Looking behind the curtain: What do L2 composition ratings really mean? *TESOL Quarterly, 29*(4), 762-765.

DeMauro, G. (1992). Examination of the relationships among TSE, TWE, and TOEFL scores. *Language Testing, 9*(2), 149-161.

Diederich, P. (1974, Spring). *Measuring growth in English*. Paper presented at the meeting of the National Council of Teachers of English, Urbana, IL.

Educational Testing Service. (1992). *TOEFL: Test of Written EnglishGuide*. Princeton, NJ: No Author.

Hamp-Lyons, L. (1995). Rating nonnative writing: The trouble with holistic scoring. *TESOL Quarterly, 29*(4), 759-762.

Kaczmarek, C. M. (1977). Scoring and rating essay tasks. In B. W. Robinette (Ed.), *Papers in ESL: Selected conferences* (pp. 151-159). Washington, DC: NAFSA.

Kroll, B. (1991). Understanding TOEFL's Test of Written English. *RELC Journal, 22*(1), 21-33.

Leki, I. (1991). A new approach to advanced ESL placement testing. *WPA: Writing Program Administration, 14*(3), 53-68.

Mohan, B., & Low, M. (1995). Collaborative teacher assessment of ESL writers: conceptual and practical issues. *TESOL Journal, 5*(1), 28-31.

Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly, 17(4)*, 651-71.

Perkins, K., & Pharis, K. (1977). TOEFL scores in relation to standardized reading tests. In B. W. Robinette (Ed.), *Papers in ESL: Selected conferences* (pp. 142-61). Washington, DC: NAFSA.

Stansfield, C. W., & Ross, J. (1988). A long-term research agenda for the Test of Written English. *Language Testing, 5*(2), 160-86.

Vaughan, C. (1991). Holistic assessment: What goes on in the rater's mind? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 111-125). Norwood, NJ: Ablex Publishing.

White, E. M. (1993). Assessing higher-order thinking and communication skills in college graduates through writing. *Journal of General Education, 42*(2), 105-122.

Authors

LYNNE DAVIS, Instructor, Center for English as a Second Language, Department of Linguistics, Southern Illinois University, Carbondale, IL 62901-4518. Email: ldavis@siu.edu. Specializations: writing instruction, spelling instruction, grammar, ellipsis in spoken English.

RUTH JOHNSON, Assistant Professor, Department of Linguistics, Southern Illinois University, Carbondale, IL 62901-4517. Email: rjohnson@siu.edu Specializations: intercultural communications, pronunciation, alternative assessment.

FLOYD OLIVE, Instructor, Center for English as a Second Language, Department of Linguistics, Southern Illinois University, Carbondale, IL 62901-4518. Email: folive@siu.edu. Specializations: L2 writing, CALL, alternative assessment.

News and Views

Applied Language Learning
1998, Vol. 9 Nos. 1 & 2, pp. 121-145

Interpreter in Action
Interview with Lieutenant Colonel Richard Francona, U.S. Air Force (Retired)

Lidia Woytak
Defense Language Institute Foreign Language Center

The star-studded Pebble Beach golf tournament was a few weeks away. It was still a sunny afternoon: Students in camouflage uniforms were rushing upstairs to the top floor auditorium. Teachers were heading for seats in the back row. Sounds of Arabic speech filled the room. Deep into the sky, one could faintly trace the sandy expanses of the desert. In some distance, a shadowy figure in an Arab garb appeared on a horse. Was it Lawrence of Arabia? Was he descending to hear his name evoked in this room?

Thanks to Mr. Bahgat Malek, a department chair of the Middle East School 2, who organized the visit of the retired Air Force Lieutenant Colonel Rick Francona, these students as well as their teachers now had the opportunity to enrich their knowledge about the job of an interpreter participating in complex multinational operations taking place in the Middle East. Francona was no stranger to DLIFLC. In 1974, he was a student of the Basic Arabic Program at the Institute and in 1978 he returned to work here as a Military Language Instructor. To this day, he considers the DLIFLC Basic Language Program superior to all other programs.

Introduced in both Arabic and English, Rick Francona emerged amidst a burst of applause to transpose the students and teachers from the safe environment of homework, tests, and exercises into the world of military operations in the Middle East. The timing for inviting a former military interpreter to DLIFLC was perfect. Today, security of the world is challenged by regional conflicts which transform regions into states and states into regions. The number of international conflicts over land, oil, and water is growing: Most of them are anticipated to continue. The United States, a United Nations (UN) member, will inevitably remain globally involved in the resolution of such conflicts.

Throughout his career, Colonel Francona was actively involved in international affairs. It began in 1976 when he participated in the

evacuation of the US Embassy in Beirut, Lebanon. Subsequently, he became an advisor to the Royal Jordanian Air Force in Amman, Jordan. Following his tour at Headquarters, U.S. Air Forces-Europe, he became a liaison officer to the Iraqi armed forces Directorate of Military Intelligence in Bagdad. Following that country's invasion of Kuwait, Colonel Francona was sent to the Gulf as the personal interpreter and advisor on Iraq to Commander-in-Chief of the U.S. Central Command, General H. Norman Schwarzkopf. In March 1991, he served as an interpreter in the Safwan cease-fire talks in Iraq.

Following a brief review of historical events in the Middle East in the recent past, Francona focused on major economic, political , and military factors defining the area. He explained that access to oil at reasonable prices is a major undercurrent of international interests. The images of Operations Desert Shield and Desert Storm were circling in listeners' minds as this regional expert drew a comprehensive picture of the events in the Gulf.

Francona's experience in the Gulf entailed a dramatic twist: From serving as an Iraqi ally during the war with Iran to becoming an opponent following its invasion of Kuwait. He described the drama of dealing with the same individuals, first as friends and then as enemies in his soon-to-be published book titled *Ally to Adversary: An Eyewitness Account of Iraq's Fall from Grace*. In the book, he guides the reader from the final Iraqi offensives in the eight-year ling Iran-Iraq war to the Iraqi defeat at the hands of the U.S. lead coalition in 1991. He also describes his experiences in dealing with the Saudis and other members of the Coalition.

During the lecture Colonel Francona frequently resorted to humor. The students burst into laughter hearing that he was not sure how to react to General Schwarzkopf's greeting, "You're the Air Force." No one could surpass his humorous description of the pompous arrival of the Iraqi generals at Safwan for the cease-fire talks. He was incredulous when the deputy chief of staff of the Iraqi Army asked him, "Which one is General Schwarzkopf?"

In many ways, Colonel Francona personified the Warrior of the 21st Century as depicted in *Joint Vision 2010:* dependable, loyal, and ready to serve. Francona did not need to describe his military readiness. He just said, "In the middle of the night I got a call telling me to be on the plane to Iraq the next day. I was." Colonel Francona advised the DLIFLC students to be prepared for an important mission ahead of time. He said, "When it comes, you have to be ready."

According to Francona, a solid language foundation and background knowledge of the area are two major elements defining a

successful career of a military linguist. During the lecture he encouraged students on several occasions to learn about the area on their own as much as they can. He added that as a student at the Institute, he found the *Area Studies Handbooks* for the Middle East very useful.

Colonel Francona also pointed out that teamwork is essential. He said that many times he was in situations in which his team members could not do the job without helping each other. Whenever the interpreters on his team needed an expert in Saudi dialect, they would turn to a colleague, for advice. Francona's personal account gave the students an overview of everyday lows and highs of the main duties of an interpreter.

Francona's presentation gave also an opportunity for the faculty to reflect upon the scope of duties of their graduates as well as their long journey from the first day in a language department to the festive graduation. Does this journey still lead to fulfilling today's needs? Do portions of the syllabus call for a readjustment? What comes first: the chicken or the egg, Modern Standard Arabic (MSA) or Iraqi?

The lecture of Colonel Francona made the listeners aware of the difficulties facing the Arabic students in the Institute's two Middle East Schools. Namely, they have to be able to select the language suitable to the text. They are tested on proficiency levels designed for one language—MSA. Yet they have to alternate between the MSA and one of several dialects to respond to a particular task. For example, Arabs would discuss current affairs in MSA, but they would conduct small talk in the Egyptian or Gulf dialects.

Notably, Richard Francona pointed out that small talk that many native speakers consider easy, is hard to master for the learner of Arabic as a second language. He stated that the unpredictability of small talk topic makes preparation for it difficult because anything may come up in an informal setting. Francona also pointed out that sometimes you cannot understand the other person without understanding some background information. While describing his interactions with the Saudis, he pointed out some speaker's tendencies to bypass what they believe to be common knowledge and focus only on new information. However, Francona amplified, what is obvious to a native speaker may not be obvious to a foreign speaker and thus may cause a breakdown in communication.

During the interview, Francona stated that teaching conversation on a one-on-one basis in his experience has been the most effective. He also recommended that students be given an opportunity to watch encounters of native speakers which could be acted out by two teachers. Such encounters, although expensive in terms of time to create, would

provide not only verbal but also kinetic models for the students to imitate. Videotaping then replaying such encounters would cut down on the expense. The basic encounters videotaped in the mid 1980s for the DLIFLC's Russian Program turned out to be successful. Natalia Goroshko and Leonid Slutsky presented their ideas on incorporating teacher encounters into curriculum in the article titled "Four-Handed Teaching." [2]

The interview with Francona makes the reader aware of complexities of skills required of an interpreter. Francona stated that while interpreting, he did not have time to think about the language. On the contrary, any time he attempted to think about the language, he found himself two sentences behind. In several instances, he alluded to the need of training in interpretation. A similar view was expressed by M. Kuwahata in "Sink or Swim: Five Basic Strokes to Consecutive Interpretation" presented during the Conference on Tradition and Innovation in Translation and Interpretation (MIIS, 1999).[1] In it she compared training in interpretation to training in swimming. Just how do we learn to swim? Is it more effective to be thrown into a deep pool so you have to somehow find your own strategies, or is it better to be taken step by step and led gradually through the various skills?, she asked. As in swimming, military linguists need training in interpreting. Daniel Gile in his book *Basic Concepts and Models for Interpreter and Translator Training* (John Benjamin, 1995) recommends activities aiming at development of comprehension abilities and verbal fluency. A completion of a rigorous six-months course in medical or court interpreting leads to a certificate. Is there a need of such a certificate for the military personnel?

During the interview, Francona stated that interpretation is an important part of military operations. Interpretation, recognized as a skill at the 1919 Paris Peace Conference and focused upon again during the famous 1949 multilingual Nuremberg trials, has also become an important part of the global language industry. In the 80's, former Congressmen Leon Panetta and Paul Simon wrote bills (H.R. 2608 and H.R. 3029) that brought interpretation and translation into the forefront of political agenda. Currently, the American Society for Testing and Materials Committee, F15.34, is reviewing national standards for language interpreting in the United States. Research on community interpretation in medical and judicial establishments indicates that not only interpreters but also personnel they work for should be aware of the basics of interpretation.[3] Perhaps the military could also apply some findings from this research.

Lieutenant Colonel Richard Francona has shown these stu-

dent-soldiers, sailors, marines, and airmen that a military linguist adopts "language for a lifetime." Although he retired from active service in 1998, he did not end his career with a retirement luncheon. He is just now beginning the most creative period of his post-service career: publishing memoirs, giving lectures, and, most important, teaching young service members and encouraging them to "be all they can be," to "aim high," and to make it an adventure and not just a job.

*
* *

Welcome, Lt. Col. Rick Francona and Major Emily Francona. It gives me great pleasure to talk to both of you. Colonel Francona, tell us about the content of your new book titled Ally to Adversary: An Eyewitness Account of Iraq's Fall from Grace *and why did you decide to write it?*

The book deals with my service as the CENTCOM interpreter in Operations Desert Shield and Desert Storm. It tells the story of my first experience in Baghdad in 1988 through a few months after the end of Desert Storm. It's a story of my observations of how Iraq changed from being a U.S. ally to one of our primary enemies in the latter half of this decade.

I wrote the book to present the history of Iraq as a personal story rather than a scholarly journal or as a one-sided correspondent account. There is always a temptation to pitimise your enemy. I wanted to tell the story, how we have come from where we were to where we are now. I wanted the book to be an easily readable, personal story–not a textbook.

To what type of readers is your book directed?

It is directed to the lay, general audience. I would imagine most people interested in it would be military people, linguists, and faculty and students from the schools here. It has very little jargon and few technical terms so the general readership can follow it. It is meant to be an entertaining, yet informative story.

Mr. Malek told me that the Iraqis did not want to deal with you initially because your language was so excellent. They thought you were a native.

They thought I might have been a Lebaneese-American. The Iraqis were

surprised that I would prefer to conduct business in Arabic. Even though one can speak Arabic, they would always provide an interpreter and begin to conduct the conversations in English. I felt more comfortable talking with them in Arabic than in English: It turned out to be about half and half. I preferred speaking in Arabic so that there would be no misunderstanding of what we were talking about.

I was concerned about their understanding of the English; and they were concerned about my understanding of the Arabic. Many times we would do it in both languages, trying to come to terms in both languages. These conversations were dealing with ongoing U.S. and Iraqi military operations. It was very important that they were done correctly.

What was your official function in Operations Desert Shield and Storm?

I was the U.S. Central Command's interpreter. We had several people, but none of us was a certified interpreter. I was responsible for getting things done. I couldn't be available 24 hours a day. I would be gone for four or five days talking to deserters or Bedouins. In the meanwhile, four or five others handled daily chores that needed to be done back in Riyahd.

I coordinated where people went, even officers senior to me. We had native speakers available in the U.S. forces, very capable officers and warrant officers. I made sure that every general officer that needed an interpreter had one available. I also made sure we had enough interpreters to rotate, because interpreting can be very nerve-wracking and intensive. We could only do it for so long before we would start loosing the edge: no more than 45 minutes without having another interpreter take it up. We would sit in the room when not interpreting, so we were not brought in cold. I don't know every Arabic word. If I didn't know a word and presumed it to be important, I would write it down and pass it behind me. We would all check each other's interpretations to make sure everything was covered and also to back each other up. It was a team effort.

How long can an interpreter perform fully in simultaneous interpretation?

After about half an hour, you start getting really, really tired because you are listening and talking almost at the same time. You never have

time to formulate what you want to say. The moment you stop to think about the meaning, you are already two sentences behind.

An interpreter's credibility is an important asset. How did you establish your credibility as an interpreter?

I already had some credibility with the Iraqis. I was known well enough to the CENTCOM staff that they asked me to come over. Then they asked me who else we needed. At that time, CENTCOM could pull anybody they wanted. I gave them four names and said, "You need these people; they are the best I know." They all performed well.

How did I establish my credibility? I established it initially with the Saudis. When we were not interpreting, we did a lot of correspondence with them. I proved to them in my daily work that I was capable of performing that task. They came to rely upon me. Thus, when there was a meeting or correspondence, they would say, "You have a good interpreter here, we don't need ours." That was how I became known. The senior Saudi staff asked for me to brief them daily in Arabic. Subsequently, I was also called on to brief the King and Minister of Defense and the King.

Could you describe the exchange of communications between the Arabic and English-speaking sides in the Gulf region?

There were very few formal negotiations or consultations with the Saudis, but there was a lot of informal verbal communication that needed interpreting: mostly informal briefings, conveyance of information, and routine exchange between the forces. Everything formal was sent via a memo or a letter. They would write to us in Arabic and we would reply in English so we were always generating our own language. The written communications we would do were fairly routine in content and vocabulary. I would do voice interpretation for the general during an informal meeting or a courtesy call. For example, the Egyptian commander would come down from his unit to visit the headquarters to pay a courtesy call on both General Khalid and General Schwarzkopf.

Could you describe the initial negotiations with the Iraqis?

The verbal exchanges stayed focused because while the translations were occurring, the principal was already thinking of the next utterance he was going to say. There was always a lag time. There was nothing said that should not have been said. Everything was thought out in advance. As we knew both the talking points and the agenda, we had a general idea of what General Schwarzkopf was going to say.

We recorded on cassette tapes formal negotiations with the Iraqis. We had four tape recorders to make sure that everything was recorded; a copy of the tapes would go to the Iraqis, and a copy of the tapes to the Coalition. There were three of us sitting at the end of the rectangular table. General Schwarzkopf and General Khalid were to my left, the Iraqi officers to my right, and the Iraqi interpreter across from me. It went in a circle. The Iraqi officers looked at each other: General Schwarzkopf and General Khalid would talk to each other. We would speak in Schwarzkopf's direction. Most of the time the opposing generals maintained eye contact with each other trying to gauge each other's reactions as soldiers do. They had time to collect their thoughts and make sure they were saying exactly what they wanted to say. It worked out well and went quickly.

Were you involved during the Second Safwan Talks?

We maintained contacts with the Iraqis almost every day during the Second Safwan Talks. I would go along for the headquarters element if a general officer was to speak. I was not the only one: There were hundreds of interpreters out there. Every American unit had its interpreters. They all did really great jobs.

One time we were interpreting a long, detailed conversation about the repatriation of 80,000 prisoners of war from Saudi Arabia back to Iraq, but this time between an Iraqi and a Saudi, both speaking Arabic. Since the U.S. officers did not understand Arabic, we were interpreting both sides of the exchange simultaneously.

We were "ganging up," as we called it. I was on one side, my colleague Vernie, on the other. I would interpret the Iraqis and he would do the Saudis to give the principal the idea of a two-way conversation. Later on, they were getting down into the specifics of how many buses per day, per hour, and what border crossings. This level of detail wasn't necessary for the U.S. general officer. Since he had a lot of other things on his mind, I gisted the conversation for him. I would say "The Iraqis are saying they need to put *x* number of buses out and the Saudis are replying that it was not enough." At the time it was sufficient to summarize the conversation for him because, later on, these conversations were going to be transcribed and typed.

Did you experience any difficulties in handling formal communications?

The more formal they got, the easier they were. Briefings in which we

conveyed significant information in a short amount of time were probably the easiest because they contained prearranged information. We had the briefing format on plastic boards and we would, in grease pencil, update the relevant information by assigning new values to the data. We gave the briefing two or three times a day in Arabic to different audiences. I could probably still give you that briefing today because they were so formalized.

Did you enjoy giving the briefings?

Yes, I did. Dealing in the military environment was probably the easiest because I was familiar with it. When I was talking to someone in a uniform, I could depend on visual clues. I could see what rank he was, so I would know what education he had, and at what level of responsibility he held. The badges would tell me what experiences he had and in what branch of service. I could adjust my speech appropriately.

Did the Saudi military use a standard memo form?

The memos from a Saudi commander or his staff were on a form. The form had its basic elements which were always the same: The date would be on the right, the serial number would be on the left, who sent it, subject, paragraph one, two, three, four, and then the signature block. It was just like a U.S. memo. The first paragraph would contain a formal greeting. However, once they got to the subject, it could be anything.

Would they get into the subject in the beginning or in the middle of the letter?

These were military people, trained in U.S. schools. They followed the U.S. standards in which the subject was followed by a statement of need. One of the most important things we translated into Arabic was the war plan. We took a lot of time to translate it because it was very detailed, specific, and critical. The Saudis made their comments on the war plan in Arabic. Afterwards we responded to it.

You encountered such speech events as negotiation, events update, discussion, address, letter, and small talk. What type of speech event was the most difficult to interpret?

Small talk. I dealt with Saudi small talk in the Gulf during Desert Shield and Desert Storm. The only small talk I would have had with any Iraqis

would have been with some shop owners when I was able to get out and move around. It was difficult as the Iraqis would neither drop their surveillance nor would they let me move around by myself—although at times I would sneak away. I had a lot of small talk with my colleague about his family, his wife and kids, and similar topics. I found talking in a dialect with the children and the wife who spent all day in the house challenging and entertaining, too. My colleague made gestures to help me understand.

Small talk was the most challenging linguistically because it could be about anything. During what I would call "down time," I was often called upon to interpret, especially during Desert Shield when we were at the headquarters waiting for the forces to arrive. They would say "Please explain to the carpenters that we need them to cut a 2x4-foot panel and to order pieces of cork." I had not often come across trade expressions. Responding to the call, I learned a lot of trade expressions that I would not have learned otherwise.

In every small talk you hear certain phrases over and over again. The more numerous and precise they are, the less capacity you need to handle them. They can free productive capacity for creative interpreting. Do you think we should train students to supply instant phrase equivalents?

Yes, rather than having to translate the phrase, they would know what the phrase is. There are a lot of those in Arabic, although, they differ from country to country. The phrases are key.

Also, abbreviations are very important and they are used more and more.

Sometimes you have to paraphrase because you don't have an equivalent.

During my lecture at DLIFLC, I told the students an anecdote about paraphrasing. After a long negotiation regarding U.S. military women driving in the Kingdom, the Saudis announced their king's decree, "American military women driving military vehicles in military uniforms are *not women*." Should we know a literal translation of that? Yes, of course. Should we use the literal translation? Obviously not. We reformulated this decree into a less offensive statement, "American military women may drive vehicles while in uniform." Same information, but not a literal translation.

Do Arabic speakers use a lot of idiomatic expressions?

Absolutely. They use many idioms in conversations, but not in formal communications. Frequently you understand the words, but the meaning makes no sense to you.

Let me share an anecdote with your readers. Whenever visiting congressmen and generals would want to buy a souvenir of Saudi Arabia, I would always be tasked to take them downtown. We would go to different merchants—they all were capable at taking money in any language. I would always tell the visitors, "I will interpret for you, but I will not negotiate for you. When this is done, you all have to agree on the price. Haggling is your job, I will merely mention the numbers." They said, "Okay, this sounds like fun." It was entertaining for them, and we would go downtown and the merchants would say, "I'm going to give you the best price." I would translate it. Then the buyer would make an offer and the merchant would reply saying, "Now listen, we're just talking bread here, not butter." And I would translate that, "He's talking bread, not butter." and I would always get this "What does that mean?" look. I would say, "I'm just telling you what he said." Then I would explain, "He's only making the bare minimum here. At that price, he would not make enough to afford butter." For a lot of these I learned standard answers, but in official business, I rarely encountered them.

What kind of expressions posed a problem for the U.S. military interpreters?

In general, technical vocabulary, military concepts, concrete terms were easy to interpret or to translate. We had problems, however, with phrases and expressions that had a cultural meaning. When the Saudi general officers were describing the morale of the forces during the war-planning discussions, they would refer to Arab history or Islamic concepts, rather than the military ones. It got confusing for us. These concepts were familiar to them because they learned them through their socialization, in school or at home. Sometimes we would get lost because we knew the words, but did not know the meaning. Other times, they would refer to another person saying, "Remember the story of *so and so*." We had never heard the story of *so and so*.

To them the story had carried a certain meaning.

Whenever we talked this way in English, it was also confusing to them. We had to decide that either this segment was important enough to stop

and ask them what they meant, or ask them if they would send it to us in writing later so that we could figure out what they meant. In their formal writing, they would not use historic or religious terms. The translation that came to us was pretty standard and stuck to the business at hand. In their free flowing conversation, however, we frequently found unfamiliar shortcuts in references to experiences that they shared.

They relied on their background.

Just as we do. We would make little quips to each other; punch lines from common jokes, stories, or English proverbs. They would do the same thing. They would recite a phrase out of the Koran which they thought conveyed the meaning. Not being conversant in the Koran, we didn't get it.

Did they sometimes use the Saudi dialect to talk in private?

No, not the Saudis. I don't recall any instances of using language as a tool to hide something from us. I know they could have. Most of the time, they were in a communicative mode. In the Command Post, we had enough understanding that if the two Saudis beside me wanted to keep their conversations confidential, they would get up and leave rather than switch to a dialect. Although many times, if they were deep into some dialect, they could have sat right there and had the conversation.

What people do with words and how they act differ from one group to another. Have you observed that the same stimulus can trigger different reactions from Americans as opposed to Middle Eastern people?

At the Institute, students learn from native Arabic instructors what is acceptable in the Middle East: what subjects are not joked about and what phrases are not accepted in common speech. Students get training here on politeness and some follow-on culture training. However, the principals do not always get this training. Sometimes they opened up a meeting with a culturally inappropriate phrase.

Frequently, officers would ask improper personal questions of the Saudi officers about their families, about their wives and children. Wives are a subject best left undiscussed; children are okay. You should know someone before you start asking personal questions, because the Saudis are private people and they do not want to discuss these subjects. Although the familiar term of address in the Arab world is to call

someone *the father of* and then give the son's name, but one should not start off with such familiarity. A lot of times the Americans would say in a friendly manner "Hi! How are you? Do you have a wife and kids?" or the really personal, and offensive, "How's your wife?" I have been trained, so I do not ask. If you do ask, the Saudis would probably reply, "Oh yes, I have a son. His name is Mahmud." Then one might think, "I'm going to be really friendly. I'm going to call this guy *Abu Mahmud*." One should not do that right off the top. When an interpreter is introducing a U.S. general to a Saudi general, and the former starts with this line of questioning, "How is your wife?" and "How are your children?" I would say "How is your family?" The Saudis know us: They know how we are, so most would not take offense anyway, but I would try to diffuse it by avoiding offensive exchanges.

Also, the subject of American politics in the Middle East is very sensitive. Most Arabs regard our support of Israel as problematic. A discussion of this subject has to be done diplomatically. Frequently we would use words and expressions that are inflammatory to an Arab. Whenever the subject came up, I would soften the translation to avoid a problem. Nowadays, our senior officers are gaining better awareness of the deep-rooted sensitivities to area politics.

Could you give an example of an inflammatory reference in politics?

A reference to Israel's right to exist citing U.N. resolutions is inflammatory. It brings up a whole range of retorts because the Arabs are well versed in all of the U.N. resolutions that the Israelis have ignored.

When you hear an utterance, you are focusing on the message. Besides the language you hear, what else do you take into consideration to formulate your message?

You look at the body language and the tone of voice. The Iraqis are like any other people; and personally I find them witty. They displayed a sense of humor even in the situation they were in. They often made quips that I found amusing. If translated, they would make no sense to the principal without an explanation that he was trying to be witty, or sarcastic, or derisive of the United States. I kept a notepad to make notes for the general. Once in a while one of the Iraqi generals would go off on a rhetorical comment about "We don't understand why the United States forces are still occupying part of our country. After all this was about Kuwait and we have left Kuwait." I would gist that for the principle. Then I would lean over on the table and write "BS" or "ranting" or "party line." So the general could look down and know this

was rhetoric; not the substance of the talk. We always had to listen to this two-minute blast of rhetoric prior to getting down to business. The Iraqis were doing it because they were told to do it, not because they believed it. For the most part, I detected no disrespect on either side: It was just something we went through. They all were initially unfriendly, but after we set up the system they, for the most part, were responding respectfully and professionally. Except one time, one Iraqi officer was very arrogant.

Why was he arrogant?

I'm not sure. We met him at Safwan. Colonel Dunn had been dealing with this general on a weekly basis. When we saw he was coming in to represent the Iraqi side, the colonel called me aside and said "This guy is always hard to deal with." I asked "Who is he?" He replied "He is the Regional Commander and takes it personally that we are occupying his country." I commented "I guess I can understand that." I made some mental notes about the Iraqi and then briefed my general. I explained that the Regional Commander was known to be hard to deal with. The general's response was "I don't care. I'm up here to get some business done and we're going to get it done." It worked out.

Language utterances manifest themselves through body language. For example, smiles of Japanese students during a lecture indicate that they are merely attentive, not necessarily agreeing. Did you occasionally get a message expressed through physical movement?

Not really, except for one occasion. We had a very long meeting with the Iraqis at the Second Safwan in which I knew the Iraqi interpreter. He was a friend of mine: In 1988 I had worked with him for months. I had been to his house, I knew his family, we had gone out to social outings together (I'm sure they were all sponsored by the Iraqi government). I knew his gestures and mannerisms. So when he was translating some rhetoric about the Iranian Revolutionary Guard Corps infiltrating Iraq, true or not, I knew this was propaganda, because he was tilting his head in a certain way. I thought, "Oh, there he goes again. He's off with the Ba'ath Party again." I could tell by his facial expressions that his heart was not in it. I looked at him, I'm sure he knew my looks and gestures as well, and gave him the "Come on. Can we speed this along?" look.

I had more awareness of Iraqi body language than Saudi. I had to learn the Saudi gestures and nods when I got there. We were there in

a cooperative atmosphere: If I did not understand something, I could easily learn it.

From the interpreter's point of view, what is the difference between the dynamics of a group versus a one-on-one talk?

Most of the time we were one-on-one. The only group settings were formal meetings or the talks with the Iraqis. I didn't really notice a change, because I was there to serve the general or senior officer present. I focused on what he wanted and what he needed; the other people in the room were his staff. Usually, we had interpreters for the U.S. officers if they needed them. At the Second Safwan Talks, however, I remember that the other U.S. officers at the table had no interpreters available and there was nothing we could do about it. Colonel Dunn and I were there to interpret for General Johnson: We gave him an ongoing account of the situation.

Would the opposing parties argue a particular meaning of a term during the negotiations?

Yes, in one instance General Schwarzkopf proposed that we establish a line on a map from which both sides would pull back by a kilometer to make a two-kilometer buffer zone. He drew a line on a map and the Iraqi interpreter used the term that meant a *political separation* rather than a *military cease-fire line*. The Iraqi general took great offense at what he thought was the United States drawing political borders inside his country. He argued with Schwarzkopf. He said, "No. I do not have the political authority to alter the political boundaries and you should not even be in my country. We can't be talking about solutions and demarcation lines." Schwarzkopf replied, "No. I'm not talking about that. I'm just talking about a line from which we can pull our soldiers back so they would not be killing each other." This exchange went back and forth for several minutes, until we finally said, "Okay, it is a line without political consequences. It is merely a line from which the military forces will withdraw." The Iraqis were concerned that the former meaning would have long-range political ramifications.

Describe the conversational style of the two opposing officers. What similarities and what differences have you observed?

The United States and the Saudi officers spoke to the Iraqis clearly, slowly, and simply. They avoided complicated terminology. General Khalid spoke to the Iraqis very distinctly in Modern Standard Arabic.

General Schwarzkopf spoke in clear, precise English.

The Iraqis, on the other hand, spoke to us in a conversational tone. In contrast to our formal, almost stilted style, the Iraqis' style was informal.

Could you describe the speech patterns of the enlisted personnel?

I found the Saudi officers easier to understand than the enlisted because of their education level. The more education you have in the Middle East, the closer your speech comes to Modern Standard Arabic. Many of the younger troops or the Bedouin that we talked to had almost no formal schooling. Iraq has a high literacy rate. Say what you will about the Saddam Hussein government, but the Ba'ath Party has raised the educational standards in both Syria and Iraq. The Iraqis are quite literate. For the most part, everybody was understandable, but occasionally you would get some less educated people. They were hard to understand, because they only knew the Iraqi dialect that they learned at home.

Saddam Hussein is getting better and better at tailoring his style to the circumstances. For example, in his speech to the nation on the Gulf crisis, he told an elaborate story from Muslim mythology. On the other hand, his conversations with his generals consisted of simple and short sentences. How important is it for an interpreter to be cognizant of different registers?

I have always had problems listening to Saddam Hussein's speech because it is very difficult. But listening to Yasir 'Araphat's speech is much easier, because he speaks in a simple style. He does not use the mythological or religious metaphors. Saddam is more of an orator who is trying to arouse Iraqi nationalism. His style is very similar to Fidel Castro's and reminiscent of Gamal Abdul Nasser who could fire up people through his rhetoric. How much information is Saddam Hussein conveying in these speeches? The best thing to do with them is to allow native Arabs who understand all the metaphors to translate them, and get the content from a transcript. When I read the transcript in English, I find that I have missed nothing because he said nothing of importance. On the other hand, when he is passing out information and decrees, he is to the point, crisp, and easy to understand.

Have you ever met Saddam Hussein?

No. The closest I got, I was in the same bar as one of his sons. We were about eight feet away and judging from the physical security that was around him and the toughness of the cadre that was with him, I decided that was the closest that I wanted to be.

You mentioned in your lecture to the DLIFLC students that there is an influence of Soviet culture on the Iraqi system. In which areas do you see similarities?

Style of government, some of the government functions, some of the intelligence services, and the security services are similar to the former Soviet system.

On the social side, the Iraqi culture in large is not that affected by the Soviets or the Russians. The Iraqi military is not organized as a carbon copy, or a smaller version of the Russian armed forces. The Iraqis have taken what they liked from the British, because the British mandate is part of their heritage, and what they got from the Russians and they have incorporated both into their own unique model. There is a mixture: They use a lot of Russian tactics and planning, but the organization of the forces is western. Iraqis have done a good job over the years of taking the best from various cultures.

Could you focus on the difference between content fidelity and linguistic fidelity? Which one do you consider more important?

Content. When I do any kind of translation or any kind of interpreting, I always strive to relay the meaning, not the words. I would always tell the general what the speaker meant to say, not the words he actually said.

There are idioms and phrases that do not translate well. For example, the *mother of all battles*, a literal translation from Arabic has became a popular phrase in English. It should be translated *greatest of all battles* or *battle of battles*; that's what it really means. The *mother of all battles* is catchy; now it is used in all sorts of contexts.

Some sentences render themselves to more than one interpretation. What made you pick one and not the other?

Context, and here we go back to what we were talking about earlier regarding expectations to discuss certain topics. I would always go with what I felt fit the context of the conversation.

How did you interpret ambiguous utterances?

I didn't. I would ask for clarification, or if it was something unimportant, I probably would let it go. If it was in an environment I was controlling, say I had an Iraqi officer, a cooperative deserter, I would ask him questions. As long as he was giving me the information that I was seeking, and he would mutter something under his breath, for example, "Well, that's life in Baghdad," I would let it go, because it wasn't germane to what I was trying to accomplish at the time. But if it was something about the subject, for example, he would point at a map and smirk, "Then there are these guys over here." I would say, "What do you mean *these guys*?" He would respond, "Don't you know?" .... and he would use a certain term. If it was different, I would say, "What do you mean by that?" He would say, "Well that's what we call the intelligence, the security guys." I said, "Tell me about them." because I didn't know. I had to decide what I was going to pursue.

In other words, you were searching for specific information.

I only got involved when there was a prisoner or a defector, or a deserter who had unique information that was pertinent to our activities. In one instance, we were trying to find a CBS journalist that had been captured by the Iraqis, Bob Simon. One of the prisoner-of-war reports, faxed in from the Saudis, referenced capturing several journalists. I said, "I need to talk to this guy."

Did you take notes during interpreting?

Yes, I wrote down almost every word I could. I tried to take down the entire conversation in my notes. Although we had the tapes, I still wanted my own notes because we had to type them up.

During your lecture at the Institute, you mentioned that sometimes you wrote down additional information for the U.S. generals.

Occasionally I did for General Johnson. At Safwan, for example, the Iraqis were responding to our expressed concern about the use of helicopter gunships against the Shi'a rebels in the south, whom we could see from our positions, yet we had given the permission to fly helicopters at the earlier meeting. In response to General Johnson's concern, the Iraqi military intelligence general went on this long diatribe of how the Iranians have infiltrated the Revolutionary Guard Corps into southern Iraq and were fomenting revolution. I would just lean over and

write, *party line*. After I gave General Johnson the gist of it, he just cut him off by saying, "It doesn't matter."

Did you finish unfinished sentences for the speaker?

No, never. I don't recall anyone ever finishing a sentence for an Iraqi or a Saudi officer. Occasionally they would say, "You know what I mean." and I would ask him to say it again. I didn't want to be in the position of trying to put someone else's thoughts, especially of an Iraqi or of a Saudi, into words unless I had their words.

Would they sometimes repeat themselves?

Yes, they would repeat utterances frequently. If they were giving an answer to a complex question, they would start, stop, and start again. During normal conversation, I would paraphrase it, or gist it, to get the meaning across.

Did they also ask you for clarifications if they could not understand what you were saying?

Yes. Once in a while, they would come to us and say "We don't understand this phrase. What did you mean?" They would have several translations of an English-language document into Arabic. They would say, "We think it means *this*" or "it could mean *this*" or "it could mean *that*" and we had to pick the one that best conveyed the meaning. That was rare, because when we wrote in English to the Saudis, and they did the same when they wrote in Arabic, we tried to be clear. I always recommended to the CENTCOM writers handling the correspondence to be to the point and to make sure it could be easily translated into Arabic. I also asked them not to use esoteric terms or abstract references. For the most part, the Saudis expressed themselves clearly to us.

Occasionally, I would take a phrase to another one of the American interpreters and say, "How would you translate this phrase? What does it mean to you" Most of the time he would respond, "It means *this* to me." I said, "I know, but in the context of the letter, this rendition does not make sense. Why would they say that?" Many times we would go to the officer that wrote it and say we were translating this text for the general. We would ask him to paraphrase it or to clear it up: That was the beauty of working with the allies. It was a little harder to clarify communications with the Iraqis.

If a speaker, an Iraqi, for example, made an error during interpretation would you let them know?

I was concerned about what their interpreter was telling the Iraqi general. I would listen to his interpretation. If it was wrong, and I thought he was getting the wrong message, I would ask to talk to the interpreter. I would try to do it in a break because I did not want to disrupt the talks. I would say "Excuse me. I think we have a problem with the translation." or, politely, "I'm not sure I understood what you said," and ask him to explain it.

Occasionally they would come to us and say, "We don't think you understood the message." The Iraqi interpreters that I dealt with were excellent speakers of English.

What would happen if you noticed an obvious mistake?

This came up during the initial talks at Safwan. The Iraqis would speak to us in Arabic and we would interpret for General Schwarzkopf into English. He would speak English and their interpreter would translate it into Arabic: always into your native language. We always checked each other's interpretation, that's one of our jobs. We had three interpreters on the U.S. side. We all noticed that the Iraqi interpreter was translating the word for *prisoners of war* and *detainees*, primarily the Kuwaiti detainees, as *guests*. We thought this was an improper interpretation, because in Arabic the word for *prisoner of war*, a military person, is very specific. They were using the right equivalent when we were talking about the U.S. and coalition prisoners, but when we were talking about the Kuwaiti civilians that had been arrested and taken to Iraq, they were using the equivalent of *guests*. We didn't disrupt the flow of the conversation, but at one of the breaks, we went over to talk to the Iraqi interpreter and we explained to him that we felt that they were using the wrong term, and we would like to correct the record. We gave him the word we preferred and they came back with a different one that we both agreed was okay. That was one of those situations where we just interjected ourselves, but we decided to intervene during a break, not the actual back-and-forth exchange between the two principals.

I noticed that Saddam Hussein also calls hostages guests *in his speeches*.

He always has. They did not use the equivalent of *hostage*. We found

a word that meant *detainees* and they agreed to use it. We didn't like the word *hostage* either.

This exchange illustrates how sensitive the use of words can be.

Since General Schwarzkopf did not understand Arabic, this improper translation of the term did not disrupt what he was doing. The senior Iraqi was not concerned about it, either. But for the historical record, we wanted it to be correct. General Khalid on the Saudi side was a little upset with that translation as well, because he understands both English and Arabic perfectly. We also told his staff that we were correcting the term and he said "Absolutely. I agree with it 100%."

This exchange was conducted in Modern Standard Arabic, not Iraqi?

Iraqi interpreter was translating from English to Arabic in a mix of "Iraqi-accented" Standard. I think too much is made of the dialects. I have found that with a solid proficiency in Modern Standard Arabic, you can go into any country and pick up enough of the local slang, vocabulary, and pronunciation to make yourself understood.

It is always great to be trained in the dialect, but sometimes we do not have that opportunity. Although I was trained at the Institute in Modern Standard with a little bit of Egyptian dialect, I never served in Egypt. We don't always have the luxury of specializing down to the dialect level in the U.S. forces. Our missions take us to Jordan, Egypt, North Africa, Saudi Arabia, and the Gulf states. From a management perspective. to train people in the Iraqi dialect only is not a good idea. Thus, Modern Standard is probably the most important tool. Modern Standard allowed me to go to any other country and pick up the dialect.

Have you experienced any lapses of attention?

Sometimes I experienced lapses of attention due to the content of a particular passage or a particular meaning. While I was figuring out meaning, he was still talking. By the time I realized the meaning, I missed part of his turn. I hoped that I could pick it up from context later.

I'm sure, occasionally, I missed things: It is just the nature of the game.

What signs of overload did you experience?

When I could not immediately pick up on what was going on. Every time that has happened to me, I have always had somebody else there.

Dunn or a Kuwaiti student that worked with us. I would just nod to them and they would pick it up.

Many of the students at the Institute would like to become interpreters. What does it take in terms of education, training, and personality to become an interpreter?

Assuming that you have mastered the required level of language, you have to understand the culture. I'm not talking about the ancient culture, arts and sciences, but about the background of these people and their recent history. If you are aware of the major issues of the Middle East, particularly following the breakup of the Ottoman Empire, you have a solid background in what has shaped the thinking of the people you are going to be dealing with. You have to like people. You cannot function as an interpreter if you are not comfortable talking to strangers at length. Also, when you are acting as an interpreter, you have to realize that you are a tool of your principal. You are not the negotiator, you are the interpreter. A lot of times there is an instinct to just cut through a problem, but you have to realize you are there as a facilitator, and not the actual conductor.

You are not supposed to step into the action, right?

It is very difficult not to do that; it is very tempting. At times, of course, everybody does it. If it is something small, it is not a problem. When General Schwarzkopf wanted to make sure there was no question about his words, he would remind me to translate exactly what he was saying.

Let us focus on training. Although job experience forges the best schools, the "tuition" can be priceless and the final outcomes irrevocable. What lessons can you transfer into schools such as DLIFLC where curriculum can be planned, time made available, and the cost controlled?

How do you train someone to the level of ancillary skills required of an interpreter? You can provide the foundation at the school. The language, of course, is the key. You have to have solid language skills to be an interpreter. There's no way around that. But that's not enough. You have to understand the history of the people you are dealing with. You have to understand also the military because you have to deal with the military situation. If you are working in a political environment, such as

an embassy, like we did in Damascus, you have to understand the political situation. You have to know your country's interests and, more importantly, your host country's interests—where they meet and where they diverge. I don't know how you replicate it in a training environment.

My experience and education prepared me for interpreting assignments. I was selected to be an interpreter in 1990, and that was the first time I did any serious interpreting. Earlier I was teaching. I had learned Arabic and worked in Arabic since graduating in 1974. I had served in Arab countries and traveled extensively in the Arab world. I had worked military and political issues in the Arab countries, and served as an adviser in a Jordanian unit. This experience gave me a good understanding of the Arab military culture.

Regional studies are very important. When I was at the Institute in 1974, we had interesting history books of each country in the language. The talk I'm giving brings that sort of thing to the students. Understanding Iraq is not reading the history of Iraq. You have to look at what happened to them. The defining characteristics of Iraq took place from the end of World War I until now. The problem with that approach is that to do it with every country that speaks Arabic would be a daunting task.

How did your training at DLIFLC prepare you for your work? What would you add or change in it to prepare today's students for their future duties?

Most of the DLIFLC students go into the Cryptologic Training System. While training and supervising graduates of the Institute, I noticed a decline of language capability in the 1980s and then an increase from the late 1980s through now. The lengthening of the program (from 47 to 63 weeks starting in 1992) had a lot to do with the improvement. I would have loved to have had 16 months of training. I say that now: I did not have to sit here for 16 months.

The schools do a good job. Even if you do not graduate speaking fluently, or understanding everything, the grammatical foundation you get in Modern Standard is good. I've served with people who have graduated from both civilian universities that have taught them Arabic and I have served with people that have gone to the State Department's Foreign Service Institute and also to contract schools. Invariably, I find that the best linguists are those that are trained at DLIFLC that have the opportunity to live in an Arabic-speaking country.

Major Emily Francona comments: Having the basics honed in country is probably the best combination you could have. I was not fortunate enough to learn any of my languages at DLIFLC. Comparing

the curriculum and the results, the Institute is definitely far superior. I would have preferred to have learned my languages at the Institute.

Any particular activities would you recommend that would help in performance of a military interpreter? You mentioned frequently you were handling routine tasks.

Translation. The interpreting was the most important task we did, but translation was the most frequent. Written documents would come to us, they were so diverse that we never knew what the subject would be. Even though they came from one military officer to another, the subject could be the *no smoking policy* in the headquarters, office assignments, communication schedules, access to prisoners, intelligence updates, or situation reports.

Could you tell us how good interpreters can prepare themselves for the first assignment?

Learn as much of the history as possible. If you are going to be called on to interpret, learn as much of the subject matter as you can, so that when you hear it, it is not a surprise to you. First, if you know the subjects that are going to come up, make sure you know the vocabulary. In one of the situations, we were going to talk about repatriation of prisoners of war. So we anticipated terms and phrases pertaining to logistics and timetables. Refresh your memory to make sure you are aware of the words the Iraqis or the Saudis may use. Second, look at a map. Familiarize yourself with the geography. Then when they mention the name of a town, you are not searching around for it on a map: You know where it is located and how it is spelled. Frequently, the principal will not only ask you for the geographical name, but also for its location. You are expected to know these basics. In summary, the more ancillary knowledge you have about the subject, the less stressful the interpreting is going to be.

In a way, you have to project into the future what will happen and what you will need when it happens. From the hind sight of your experience, what else would you do in preparation for interpreting?

I would have, probably, learned more of the Saudi dialect. When I went to Saudi Arabia, I knew I was going to be General Schwarzkopf's interpreter. I assumed correctly that I would be talking to Iraqis or debrief-

ing Iraqi prisoners of war. The majority of the time, however, I talked to the Saudis in Modern Standard. Although it was adequate, I wish I could have communicated with them in their dialect.

There are numerous training aids such as dialect textbooks. You say "I'm going to Saudi Arabia." Do I need Saudi dialect? Do I need Iraqi dialect? Who am I going to be talking to most? I felt comfortable with the Iraqis and the Saudis, but I could have been a little more effective, on an interpersonal basis, had I spoken a little more Saudi dialect. We had one interpreter assigned to us, an army lieutenant colonel reservist, who had gone to the Saudi Command College and was just wonderful in Saudi dialect. A little slow in the Iraqi dialect, but great in Saudi. Because of his proficiency in this dialect, his rapport with the Saudis was great. I found that rapport is important if you are going to be interpreting for the same people all the time, or briefing the same people. Good rapport makes things easier.

In other words, you have to, partially, fit a person to the task.

Absolutely. We rarely talked to the same Iraqi twice. We would do our job, and we probably would never see them again. In the headquarters, I saw the same Saudi officers everyday; twice a day sometimes. For example, I got to know General Madani on a social and professional level because I briefed him in Arabic twice a day.

What enhancement routines, that you have used, would you recommend for our graduates?

I kept current with radio and print media. Today, Arabic satellite, cable programming, an expanded VTT (Video Tele Training) system and internet content, provide excellent sources for language enhancement. All these tools have great potential for one-on-one tailored training, as well as language maintenance and enhancement.

Thank you very much for your advice to our students. I wish you and your wife success in your future endeavors.

Notes

[1] The Conference on *Tradition and Innovation in Translation and Interpretation* took place in February 1999 at the Monterey Institute of International Studies.

[2] N. Goroshko & L. Slutsky. (1993). Four-Handed Teaching. *Dialog on Language Instruction, 9*/1, 49-53.

[3] H. Mikkelson. (1998). Towards a Redefinition of the Role of the Court Interpreter. *Interpreting: International Journal of Research and Practice in Interpreting, 3/2 42.*

Reviews

Teaching by Chatting. (1998). By J. M. Cots. Lleida: Universitat de Lleida. Pp. 205, paper, ISBN 84-89727-66-X.

Reviewed by LEO VAN LIER
Monterey Institute of International Studies

This book, the first monograph in a series called *Quadernas de Sintagma* (*Sintagma* is a journal published by the University of Lleida), is subtitled "A Pragmatic Analysis of Instructor-Student Conversations at an American University." It provides a detailed discussion of what is called "office hours" at universities, a topic that has not been extensively investigated, even though it provides an important academic context for university students. The corpus of data consists of 20 office-hour conversations, ranging from less than two minutes to over 40 minutes.

The study is an interesting exercise in the analysis of spoken interaction from several different perspectives. But those who are looking for explicit advice on the office hour, or for a critical perspective on the professor–student relationship as developing outside of lectures and classrooms, will be disappointed. The study is an explicitly neutral discussion of office-hour discourse, and the reader will have to decide what to think of it in academic and pedagogical terms. The analysis is like a mirror: Ultimately readers have to decide what it is that they see, and how they like it.

The first part of the book is a general introduction to communicative and pragmatic competence. Apart from a brief introductory "walking tour" in which role plays are used to show how non-native and native speakers participate in academic communicative encounters, this section is not explicitly related to the topic of office hours. Rather, it is an overview of various theories and models of spoken interaction. This section, consisting of about fifty pages, is a very lucid and well-organized overview of various models of communicative competence and of spoken interaction. As an introduction to these topics, it is competently and efficiently done, and worth reading for its own sake, quite apart from any interest the reader may have in the office-hour encounter itself.

The second and main part of the book consists of an

ethnographic description of instructor-student interaction during office hours. In this description the various models and theories from the first part of the book are used to elucidate samples of instructor–student interaction. As Cots points out, the office hour encounter falls between the institutional and the personal areas of social action, more so than the classroom or lecture hall.

Cots uses the notion of self-presentation to illustrate strategies such as "avoiding assertiveness," "explaining and justifying," "displaying a positive self," "emphasizing modesty," and "showing attitudes and feelings." He also discusses the negotiation of formality, power and distance, and the minimization of imposition.

A separate chapter deals with discourse competence by analyzing topic management, turn taking, and information structure. Repairing is discussed under the heading of strategic competence, although the boundary between discourse competence and strategic competence seems rather blurred. The final chapter shows the various aspects of analysis in one coherent whole by analyzing one complete encounter (albeit a short one) using all the categories and strategies explained in the book. The appendix contains transcripts of three further encounters that could be used by readers or students of conversation analysis to conduct similar analyses.

The book contains a thorough overview of the analysis of spoken discourse in the context on one type of speech event, the office hour encounter at an American university. It integrates four approaches to descriptive pragmatics: sociolinguistics, conversation analysis, discourse analysis, and the ethnography of speaking. It serves as a very useful text on descriptive pragmatics and sheds interesting light on an important but not much studied speech event. It deserves to be widely read by students of sociolinguistics and pragmatics, as well as by those who hold office hours.

Language and Development: Teachers in a Changing World. (1997). Editors: B. Kenny and W. Savage. (*Applied Linguistics and Language Study Series*. General Editor: *Christopher N. Candlin*.) New York: Addison Wesley Longman.*

Reviewed By KEVIN W. K. CHU
City University of Hong Kong

This collection of selected papers was originally presented at the Conference of the Regional Education Language Center (RELC) on Language Programs in Development Projects in April 1993 at the Asian Institute of Technology in Bangkok. The papers document English Language Teaching in Cambodia, China, India, Indonesia, Laos, Malaysia, Papua, New Guinea, the Philippines, Singapore, and Vietnam. The book illuminates "short stories of teachers' professional lives in a changing world" rather than scholarly written papers (p. 4). The authors of **Language and Development: Teachers in a Changing World** (hereafter **L and D**) discuss recent education policy, English for Special Purposes (ESP) issues, and foreign aid for foreign language education in these countries. The book is organized into three parts titled *Coping with Change*, *Teaching and Learning in Different Worlds*, and *Responding to the Players*; and consists of 21 chapters plus introductory and concluding chapters.

The first seven chapters in **L and D** demonstrate how practitioners "cope with change" by reconstructing education programs, introducing staff development programs, by developing teacher support teams, identifying students' purposes of learning English, and by using task-based activities such as modified role play. The papers focus on practical concerns such as large-class size, low motivation, lack of budget and staff, and time constraints.

Teaching and Learning in Different Worlds is also composed of seven chapters which document cases of adapting culturally incompatible teaching materials and also cases of integrating language teaching with mainstream subjects in the forms of simulation exercises and business visit projects. In order to implement teacher education reforms, either in classes of trainee pilots, engineering students or electrical technicians, language teachers, educators, and development professionals struggle with traditional beliefs, bureaucracy, and fear of change. The chapters exemplify well how reflective the practitioners are in making these adaptations to "different worlds."

The area of the conflicting objectives and needs of the foreign language education programs is discussed in *Responding to the Play*

ers. By the *Players*, Kenny and Savage refer to those who deliver and receive foreign aid for foreign language education. Some telling stories in this section will initiate discussion of the roles and expectations of funding agencies, managing agencies, language teaching consultants, politicians, academics, and teachers.

I appreciated the clear account of characteristics in the concluding chapter of **L and D**. Here, Savage admirably justifies the new emerging field of language and development by arriving at its five notable characteristics. Namely, they are: (1) change-oriented, (2) experiential, (3) pro-autonomous, (4) collaborative, and (5) communicative. This interpretative commentary is a good beginning of a discussion forum for language planning issues, linguistic imperialism, and language rights that have not been made explicit in these "developing" societies. I hope that future language-and-development experts will expand on these issues at length. Nonetheless, **L and D** is extremely valuable in presenting the development of language training; in establishing the forum for the voices of the local teachers and expatriate language educators; in raising awareness of the effect of the cultural, economic and sociopolitical factors on language policies; and, finally, in providing practical suggestions for teachers. In this regard, the stories present the readers with concrete images of how to implement a critical pedagogy—how efficient the players might be in applying professional knowledge; how sensitive to the institutional and social contexts they ought to be; and how, finally, they could not simply accept things as they were.

We can see the enthusiasm of **L and D** to show us around "the changing world." The chapters can be read properly only in the light of full appreciation of the difficulties confronted by the players, the frontline teachers and, most importantly, the course participants. I have no doubts that readers could arrive at their own meanings, as invited by Kenny and Savage in the introductory chapter titled "Setting the Scene," and could reflect upon the experience shared by the contributors. How are their practices constrained and influenced by institutional, social, historical, ethical, and political factors? By unveiling a wider scope of teaching contexts, this book helps readers question taken-for-granted and unarticulated assumptions and consequently alter their perspectives on language study and language education.

*This is a substantially expanded version of the book review published in TESL-HK, a newsletter for ELT professionals in Hong Kong, June 1998, 2.

Making Communicative Language Teaching Happen: Directions for Language Learning and Teaching. (1995). By James F. Lee and Bill VanPatten. New York: McGraw-Hill.

Reviewed by PHILIP A. WHITE
Defense Language Institute Foreign Language Center

Many teachers of foreign languages may have come to the field from different ethnic and professional groups, Some learned the language they teach at an institution, while others grew up with it at home. Especially the latter may want to gain a better understanding of the field of foreign language education and contemporary methods. If so, then this book, *Making Communicative Language Teaching Happen*, could served as a good start.

As the title indicates, this text provides a basic introduction to the approach of communicative language teaching. The two authors are well prepared to do so as both are widely known within the field of foreign language education. Thus, the reader should have some assurance of the usefulness and "authenticity" of the material.

The authors have written this text as a "guide to helping instructors develop a communicative classroom environment that blends listening, speaking, and writing (p. ix)." Their audience is those who are graduate students (the book is planned as a text for a course having been "field tested" with the authors' students) and practicing teachers who need a resource manual for developing tasks and material for the classroom. They note that there over 200 activities and test sections included (p. x).

Topics include new roles for teachers and students, the importance of "comprehensible mean-bearing input in second language acquisition," an approach to grammar instruction, classroom oral communication with suggestions, contemporary approaches to teaching reading and writing, suggestions on building towards proficiency, and testing principles. The book contains suggestions and questions for reflection throughout the book, suggestions for further reading, and end-of-chapter activities (presumably for classroom and students using the text in a course), and an associated workbook for the text (not reviewed here).

The book contains 13 chapters and an epilogue, organized into five sections of varying size, a bibliography, and an index. The five major sections of the text are: (1) Preliminary considerations, (2) Grammar instruction, (3) Spoken language, (4) Reading and writing and (5) A look forward.

The material the authors provide is contemporary and reflective of the field as well as indicative of the approach they use. In some areas perhaps the reader is offered too much material that might be of minimal interest to the teacher or persons now in preparation for teaching. Particularly their discussion of the older, and now generally unused audiolingual method, is more defensive in nature than informative and useful for the classroom (p. 7).

After establishing a view of a classroom teacher that the two authors consider common in teaching, that of the instructor as an "Atlas figure" holding the entire weight of the classroom and learners on his shoulders, they offer a countervailing perception of the teacher as an architect and resource person within the class (chapters 1 and 2). Unfortunately, they then proceed to equate the instructor-as-Atlas view with audiolingualism. While the view of the teacher as a facilitor and as a resource has much to recommend it, the defensive attacks on audiolingualism neither offer much to the reader nor enhance the text.

Lee and VanPatten provide a detailed and useful view of teaching grammar within the communicative language teaching approach. The debate over the place of grammar in the foreign language classroom has been a historic one within the field and is not limited to any one approach. The authors come down on the side of teaching grammar and doing so through "structured" input and output in the classroom. Rather than have learners perform activities that force manipulation of their output, the view offered here is that of structured input to offer the learner an opportunity for meaning-bearing activities. In support of this approach to grammar, they offer a wealth of activities for input and output (chapters 5 and 6).

Readers who find the presentation of grammar at least somewhat alien to a traditional view should be aware that the two authors are admittedly influenced by the work of S. Pit Corder, a British educator who is closely associated with the concept of "code switching" within the field of sociolinguistics as well as error analysis. At the beginning, they cite Corder's view that language might be an activity that cannot be taught, but can only come about within conditions conducive to acquisition by the learner (p. 35).

Readers who are seeking a contemporary means of handling pronunciation in the classroom will be disappointed by the writings. While the text contains a discussion of listening comprehension with appropriate exercises, pronunciation does not even appear in the index. As most people, whether teachers or students, are aware that normally human language involves the production of sound and that the sounds of another language are rarely identical to their native speech,

the absence of any discussion of pronunciation is puzzling. In a text claimed as being developed for practicing teachers, some sort of statement about pronunciation seems obviously necessary—why are or aren't activities specifically for pronunciation necessary? How is the learner supposed to make him or herself comprehensible to the native speaker? On these questions as well, the authors are silent.

As noted, the text contains many activities for the classroom. From the perspective of those who teach languages located outside of western Europe, the restriction of examples to primarily Spanish could be viewed as a limiting factor. Of course, as the authors hold positions in university Spanish departments this limitation perhaps is understandable.

All in all, in spite of its faults—and what book is flawless?—**Making Communicative Language Teaching Happen** is a text that can be read with much profit by most classroom teachers. Most readers will benefit not only from the example activities, but also from the extensive suggestions for reading given at the end of each chapter.

The Road Ahead. (1995). By William H. "Bill" Gates, III. New York: Viking. Pp. 286, CD version included, ISBN 0-670-77289-5.

Reviewed By RODERIC A. GALE
Defense Language Institute Foreign Language Center

"First and foremost, Bill Gates is an idea man," said Barbara Walters on December 22, 1998 during the ABC News Special *The 10 Most Fascinating People of 1998*. This characterization clearly reflects the message coming from *The Road Ahead*.

In keeping with everything one expects of the Microsoft Mastermind, this book is always looking forward, with just enough history to set the context for that look and to prove the idea that looking forward is the only way to go, whether in dealing with information systems or anything else. Selections from the Contents page give insight to that direction: A Revolution Begins, Lessons from the Computer Industry, Applications and Appliances, Implications for Business, Education: The Best Investment, and Critical Issues.

In reviewing this book for the *Applied Language Learning*, I have looked at it from the perspective of what it has to offer for those involved in language teaching and evaluation. However, before addressing that assessment, an overview of the book is warranted. Bill Gates did not write this, his first publication, as a historical review of his life, of the software business, or his contributions to that business. Those factors do appear throughout the book, but only to provide the setting to look forward and the means by which events have moved forward. It is interesting that the perhaps richest person in the world has not focused on what he has done, but rather on what has been done and the lessons to be drawn from his work. His thoughts focus on what the future holds in store and how we can either seize the opportunities ahead or be held captive to them.

Just as he speaks in public or television appearances, Gates writes with a casual approach that makes for a fast read. It is not at all what one may expect from someone who virtually created the high tech software world which impacts everyone around the world in one way or another. There is no technical information to be gleaned and very few notes to be taken on the ideas he presents. As a result, it can be read in a weekend or over a week of evenings. The compartmentalization of the chapters makes it easy to move through the book at will.

What of the chapter that most applies to language education though? "Some fear that technology will dehumanize the formal

educator. . . . But. . . technology can humanize the education environment. Corporations (read, teaching organizations) are reinventing themselves around the flexible opportunities afforded by information technology, classrooms will have to change as well. (p. 184) There is an often-expressed fear that technology will replace teachers. I can say emphatically and unequivocally, IT WON'T. However, technology will be pivotal in the future role of teachers" (p. 185). I believe one can say the same for testers and evaluators of students. As a result, the full spectrum of people in the foreign language process—students, teachers, and testers—should be challenged and assured by his comments.

Clearly, this man Bill Gates, who may arguably be the person most influential in putting computer software in the home and classroom, is calling for those in education to leverage the use of computers in their teaching. He adamantly rejects the thought that machines and programs will replace these educators. "Educators. . . are, among other things, facilitators. . . they will have to adapt and readapt to changing conditions. Unlike some professions, however, the future of teaching looks extremely bright" (p. 187).

The Road Ahead chapters on Education: The Best Investment and Critical Issues are part of the spectrum of light along the tunnel of education. We in that profession may do well to rally behind this White Knight.

Calendar of Events*

1999

6-9 March, *American Association of Applied Linguistics*, Stam ford. Information AAAL, (612) 953-0805, Fax (612) 431-8404, PO Box 21686, Eagan, MN 55121-0686; Email [aaaloffice@aaal.org].

8-14 March, *Teachers of English to Speakers of Other Lan guages*, New York. Information TESOL, (703) 836-0774, Fax (703) 836-7864, 1600 Cameron St., Suite 300, Alexan dria, VA 22314-2751; Email [conv@tesol.edu], URL [www.tesol.edu].

11-13 March, *Southern Conference on Language Teaching with Foreign Language Association of Virginia*, Virginia Beach. Information Lynne McClendon, SCOLT Executive Director, (770) 992-1256, 165 Lazy Laurel Chase, Roswell, GA 30076; Email [lynnemcc@mindspring.com].

7-10 April, *Pacific Northwest Council for Languages*, Tacoma. Information PNCFL, PO Box 4649, Portland, OR 97208-4649; Email [112063.622@compuserve.com].

8-11 April, *Northeast Conference on the Teaching of Foreign Languages*, New York. Information Northeast Conference, Dickinson College, (717) 245-1977, Fax (717) 245-1976, PO Box 1773, Carlisle, PA 17013-2896; Email [nectfl@dickinson.edu], URL [www.dickinson.edu/nectfl].

8-11 April, *American Hungarian Educators Association*, Cleve land. Information Martha Pereszlenyi-Pinter, Classical & Modern Languages & Cultures, (216) 397-4723, FAX (216) 397-4256, John Carroll University, Cleveland, Ohio 44118; Email [mpereszlenyi@jcvaxa.jcu.edu], URL [http://www.magyar.org/home.html].

15-18 April, *Central States Conference on the Teaching of Foreign Languages*, Little Rock. Information CSCTFL, Rosalie Cheatham, (501) 569-8159, Fax (501) 569-8157, University of Arkansas - Little Rock, 2801 S. University Avenue, Little Rock, AR 72204; Email [rmcheatham@ualr.edu].

*Courtesy of *The Modern Language Journal* (University of Wisconsin)

13-15 May, *JNCL-NCLIS Delegate Assembly*, Washington. Information JNCL-NCLIS, (202) 966-8477, 4646 40th St. NW, Third Floor, Washington DC 20016; Email [info@languagepolicy.org], URL [http://www.languagepolicy.org].

20-23 May, *Language Teacher Education*, Minneapolis. Information International Conference on Language Teacher Education, CARLA, (612) 627-1870, Fax (612) 624-1875, UTEC, Suite 111, 1313 5th St SE, Minneapolis, MN 55414; E-mail [carla@tc.umn.edu], URL [http://carla.acad.umn.edu].

22-30 May, *Conseil International d'Etudes Francophones*, Lafayette. Information Ginette Adamson, Fax (316) 978-3319, Modern Languages, Wichita State University, Wichita, KS 67260-0011; Email [adamson@twsuvm.uc.twsu.edu].

30 May-4 June, *Computer Assisted Language Instruction Consortium*, Oxford. Information Esther Callais, (512) 245-1417, Department of Modern Languages, Southwest Texas State University, San Marcos, TX 78666; Email [info@calico.org].

3-6 June, *ADFL Seminar West*, Palo Alto. Information Association of Departments of Foreign Languages, Attn: Elizabeth Welles, 10 Astor Place, New York, NY 10003-6981; Email [elizabeth.welles@mla.org].

11-14 July, *American Association of Teachers of French*, St. Louis. Information AATF, (618) 453-5731, Fax (618) 453-5733, Mailcode 4510, Department of Foreign Languages, Southern Illinois University, Carbondale, IL 62901-4510; Email [abrate@siu.edu], URL [aatf.utsa.edu].

30 July-3 August, *American Association of Teachers of Spanish & Portuguese*, Denver. Information AATSP, (970) 351-1090, Fax (970) 351-1095, Butler-Hancock Hall #210, University of Northern Colorado, Greeley, CO 80639; Email [lsandste@bentley.unco.edu].

16-21 August, *International Association of Teachers of Russian Language and Literature*, Bratislava (Slovakia). Information American Council of Teachers of Russian (ACTR), (202)833-7522, Fax (202) 833-7523, 1776 Massachusetts Ave. NW, Suite 700, Washington, DC 20036; Email [ddavidson@actr.org].

28-30 October, *Foreign Language Association of North Carolina*, High Point. Information Debra S. Martin, FLANC Executive

Director, (828) 686-4985, Fax (828) 686-3600, PO Box 19153, Asheville, NC 28815; Email [martintl@interpath.com].

4-6 November, *Wisconsin Association of Foreign Language Teachers*, Appleton. Information Kyle Gorden, (414) 723-6316, 4969 Hickory Court, Elkhorn, WI 53121; Email [kylegorden@elknet.net].

17-18 November, *National Association of District Supervisors of Foreign Languages*, Dallas. Information Sharon Watts, (402) 557-2440, Omaha Public Schools, 3215 Cuming, Omaha, NE 63131; Email [swatts@ops.org].

18-21 November, *American Association for the Advancement of Slavic Studies*, St. Louis. Information AAASS; Email [walker@core-mail.fas.harvard.edu].

19-21 November, *American Council on the Teaching of Foreign Languages*, Dallas. Information ACTFL, (914) 963-8830, Fax (914) 963-1275, 6 Executive Plaza, Yonkers, NY 10701-6801; Email [actflhq@aol.com], URL [http://www.actfl.org].

19-21 November, *American Association of Teachers of German*, Dallas. Information AATG, (609) 795-5553, Fax (609) 795-9398, 112 Haddontowne Court #104, Cherry Hill, NJ 08034; Email [73740.3231@compuserve.com].

27-30 December, *Modern Language Association of America*, Chicago. Information MLA, Fax (212) 477-9863, 10 Astor Place, New York, NY 10003-6981; Email [convention@mla.org].

27-30 December, *North American Association of Teachers of Czech*, Chicago. Information Masako Ueda, (401) 863-3933, Fax (401) 863-7330, Box E, Department of Slavic Lan guages, Brown University, Providence, RI 02912; Email [masako_ueda@brown.edu].

27-30 December, *American Association of Teachers of Slavic & E. European Languages*, Chicago. Information AATSEEL, Fax (520) 885-2663, 1933 N. Fountain Park Dr., Tucson, AZ 85715; Email [76703.2063@compuserve.com], URL [http://clover.slavic.pitt.edu/~aatseel/].

2000

24-26 February, *Southern Conference on Language Teaching with Alabama Association of Foreign Language Teachers*,

Birmingham. Information Lynne McClendon, SCOLT Executive Director, (770) 992-1256, 165 Lazy Laurel Chase, Roswell GA 30076; Email [lynnemcc@mindspring.com].

10-13 March, *Central States Conference on the Teaching of Foreign Languages*, TBA. Information CSCTFL, Rosalie Cheatham, (501) 569-8159, Fax (501) 569-8157, University of Arkansas - Little Rock, 2801 S. University Avenue, Little Rock, AR 72204; Email [rmcheatham@ualr.edu].

11-14 March, *American Association of Applied Linguistics*, Vancouver. Information AAAL, (612) 953-0805, Fax (612) 431-8404, PO Box 21686, Eagan, MN 55121- 0686; Email [aaaloffice@aaal.org].

14-18 March, *Teachers of English to Speakers of Other Lan guages*, Vancouver. Information TESOL, (703) 836-0774, Fax (703) 836-7864, 1600 Cameron St., Suite 300, Alexan dria, VA 22314-2751; Email [conv@tesol.edu], URL [www.tesol.edu].

13-15 April, *Pacific Northwest Council for Languages*, Missoula. Information PNCFL, PO Box 4649, Portland, OR 97208-4649; Email [112063.622@compuserve.com].

13-16 April, *Northeast Conference on the Teaching of Foreign Languages*, Washington DC. Information Northeast Con ference, (717) 245-1977, Fax (717) 245-1976, Dickinson College, PO Box 1773, Carlisle, PA 17013-2896; Email [nectfl@dickinson.edu], URL [www.dickinson.edu/nectfl].

4-6 May, *Balkan and South Slavic Linguistics, Literature and Folkore*, Lawrence. Information Marc L. Greenberg, Dept. of Slavic Languages and Literatures, Fax (785) 864-4298, 2134 Wescoe Hall, Lawrence, KS 66045-2174; Email [m-greenberg@ukans.edu].

TBA July, *American Association of Teachers of French*, Paris. Information AATF, (618) 453-5731, Fax (618) 453-5733, Mailcode 4510, Department of Foreign Languages, Southern Illinois University, Carbondale, IL 62901-4510; Email [abrate@siu.edu].

1-5 August, *American Association of Teachers of Spanish & Portuguese*, San Juan. Information AATSP, (970) 351-1090, Fax (970) 351-1095, Butler-Hancock Hall #210, University of Northern Colorado, Greeley, CO 80639; Email [lsandste@bentley.unco.edu].

2-4 November, *Foreign Language Association of North Carolina*, High Point. Information Debra S. Martin, FLANC Executive

Director, (828) 686-4985, Fax (828) 686-3600, PO Box 19153, Asheville, NC 28815; Email [martintl@interpath.com].

9-12 November, *American Association for the Advancement of Slavic Studies*, Denver. Information AAASS; Email [walker@core-mail.fas.harvard.edu].

15-16 November, *National Association of District Supervisors of ` Foreign Languages*, Boston. Information Sharon Watts, (402) 557-2440, Omaha Public Schools, 3215 Cuming, Omaha, NE 63131; Email [swatts@ops.org].

17-19 November, *American Council on the Teaching of Foreign Languages*, Boston. Information ACTFL, (914) 963-8830, Fax (914) 963-1275, 6 Executive Plaza, Yonkers, NY 10701-6801; Email [actflhq@aol.com], URL [http:// www.actfl.org].

17-19 November, *American Association of Teachers of German*, Boston. Information AATG, (609) 795-5553, Fax (609) 795-9398, 112 Haddontowne Court #104, Cherry Hill, NJ 08034; Email [73740.3231@compuserve.com].

27-30 December, *Modern Language Association of America*, Washington, D.C. Information MLA, Fax (212) 477-9863, 10 Astor Place, New York, NY 10003-6981; Email [convention@mla.org].

27-30 December, *North American Association of Teachers of Czech*, Washington, D.C. Information Masako Ueda, (401) 863-3933, Fax (401) 863-7330, Box E, Department of Slavic Languages, Brown University, Providence, RI 02912; Email [masako_ueda@brown.edu].

27-30 December, *American Association of Teachers of Slavic & E. European Languages*, Washington, D.C. Information AATSEEL, Fax (520)885-2663, 1933 N. Fountain Park Dr., Tucson, AZ 85715; Email [76703.2063@compuserve.com], URL [http://clover.slavic.pitt.edu/~aatseel/].

2001

24-27 February, *American Association of Applied Linguistics*, St. Louis. Information AAAL, (612) 953-0805, Fax (612) 431-8404, PO Box 21686, Eagan, MN 55121-0686; Email [aaaloffice@aaal.org].

27 February-3 March, *Teachers of English to Speakers of Other Languages*, St. Louis. Information TESOL, (703) 836-0774,

Fax (703) 836-7864, 1600 Cameron St., Suite 300, Alexandria, VA 22314-2751; Email [conv@tesol.edu], URL [www.tesol.edu].

19-22 April, *Northeast Conference on the Teaching of Foreign Languages*, New York. Information Northeast Conference, (717) 245-1977, Fax (717) 245-1976, Dickinson College, PO Box 1773, Carlisle, PA 17013-2896; Email [nectfl@dickinson.edu], URL [www.dickinson.edu/nectfl].

14-15 November, *National Association of District Supervisors of Foreign Languages*, Washington, D.C. Information Sharon Watts, (402) 557-2440, Omaha Public Schools, 3215 Cuming, Omaha, NE 63131; Email [swatts@ops.org].

16-18 November, *American Council on the Teaching of Foreign Languages*, Washington, D.C. Information ACTFL, (914) 963-8830, Fax (914) 963-1275, 6 Executive Plaza, Yonkers, NY 10701-6801 Email [actflhq@aol.com], URL [http://www.actfl.org].

16-18 November, *American Association of Teachers of German*, Washington, D.C. Information AATG, (609) 795-5553, Fax (609) 795-9398, 112 Haddontowne Court #104, Cherry Hill, NJ 08034; Email [73740.3231@compuserve.com].

27-30 December, *Modern Language Association of America*, TBA. Information MLA, Fax (212) 477-9863, 10 Astor Place, New York, NY 10003-6981; Email Convention@mla.org].

27-30 December, *North American Association of Teachers of Czech*, TBA. Information Masako Ueda, (401) 863-3933, Fax (401) 863-7330, Box E, Department of Slavic Lan guages, Brown University, Providence, RI 02912; Email [masako_ueda@brown.edu].

27-30 December, *American Association of Teachers of Slavic & E. European Languages*, TBA. Information AATSEEL, Fax (520) 885-2663, 1933 N. Fountain Park Dr., Tucson, AZ 85715; Email [76703.2063@compuserve.com], URL [http://clover.slavic.pitt.edu/~aatseel/].

Information for Contributors

Statement of Purpose

The purpose of *Applied Language Learning* (*ALL*) is to increase and promote professional communication within the Defense Language Program and academic communities on adult language learning for functional purposes.

Submission of Manuscripts

The Editor encourages the submission of research and review manuscripts from such disciplines as: (1) instructional methods and techniques; (2) curriculum and materials development; (3) testing and evaluation; (4) implications and applications of research from related fields such as linguistics, education, communication, psychology, and social sciences; (5) assessment of needs within the profession.

Research Article

Divide your manuscript into the following sections:

- Abstract
- Introduction
- Method
- Results
- Discussion
- Conclusion
- Appendices
- Notes
- References
- Acknowledgements
- Author

Abstract
Identify the purpose of the article, provide an overview of the content, and suggest findings in an abstract of not more than 200 words.

Introduction
In a few paragraphs, state the purpose of the study and relate it to the hypothesis and the experimental design. Point out the theoretical implications of the study and relate them to previous work in the area.

Next, under the subsection *Literature Review,* discuss work that had a direct impact on your study. Cite only research pertinent to a specific issue and avoid references with only tangential or general significance. Emphasize pertinent findings and relevant methodological issues. Provide the logical continuity between previous and present work. Whenever appropriate, treat controversial issues fairly. You may state that certain studies support one conclusion and others challenge or contradict it.

Method

Describe how you conducted the study. Give a brief synopsis of the method. Next develop the subsections pertaining to the *participants,* the *materials,* and the *procedure.*

Participants. Identify the number and type of participants. Specify how they were selected and how many participated in each experiment. Provide major demographic characteristics such as age, sex, geographic location, and institutional affiliation. Identify the number of experiment dropouts and the reasons they did not continue.

Materials. Describe briefly the materials used and their function in the experiment.

Procedure. Describe each step in the conduct of the research. Include the instructions to the participants, the formation of the groups, and the specific experimental manipulations.

Results

First state the results. Next describe them in sufficient detail to justify the findings. Mention all relevant results, including those that run counter to the hypothesis.

Tables and figures. Prepare tables to present exact values. Use tables sparingly. Sometimes you can present data more efficiently in a few sentences than in a table. Avoid developing tables for information already presented in other places. Prepare figures to illustrate key interactions, major interdependencies, and general comparisons. Indicate to the reader what to look for in tables and figures.

Discussion

Express your support or nonsupport for the original hypothesis. Next examine, interpret, and qualify the results and draw inferences from them. Do not repeat old statements: Create new statements that further contribute to your position and to readers understanding of it.

Conclusion

Succinctly describe the contribution of the study to the field. State how it has helped to resolve the original problem. Identify conclusions and theoretical implications that can be drawn from your study.

Appendices
Place detailed information (for example, a table, lists of words, or a sample of a questionnaire) that would be distracting to read in the main body of the article in the appendices.

Notes
Use them for substantive information only, and number them serially throughout the manuscript. They all should be listed on a separate page entitled *Notes*.

References
Submit on a separate page of the manuscript a list of references with the centered heading: *References*. Arrange the entries alphabetically by surname of authors. Review the format for bibliographic entries of references in the following sample:

> Dulay, H., & Burt, M. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly, 16* (1), 93-95.
> Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill.

List all works cited in the manuscripts in *References*, and conversely, cite all works included in *References* in the manuscript. Include in reference citations in the text of the manuscript the name of the author of the work cited, the date of the work, and when quoting, the page numbers on which the materials that you are quoting originally appeared, e.g., (Jones, 1982, pp. 235-238).

Acknowledgments
Identify colleagues who contributed to the study and assisted you in the writing process.

Author
Type the title of the article and the author's name on a separate page to ensure anonymity in the review process. Prepare an autobiographical note indicating: full name, position, department, institution, mailing address, and specialization(s). Example follows:

> JANE C. DOE, Assistant Professor, Foreign Language Education, University of America, 226 N. Madison St, Madison, WI 55306. Specializations: foreign language acquisition, curriculum studies.

Review Article

It should describe, discuss, and evaluate several publications that fall into a topical category in foreign language education. The relative significance of the publications in the context of teaching realms should be pointed out. A review article should be 15 to 20

double-spaced pages.

Review

Submit reviews of textbooks, scholarly works on foreign language education, dictionaries, tests, computer software, video tapes, and other non-print materials. Point out both positive and negative aspects of the work(s) being considered. In the three to five double-spaced pages of the manuscript, give a clear but brief statement of the work's content and a critical assessment of its contribution to the profession. Keep quotations short. Do not send reviews that are merely descriptive.

Manuscripts are accepted for consideration with the understanding that they are original material and are not being considered for publication elsewhere.

Specifications for Manuscripts

All editorial correspondence, including manuscripts for publication should be sent to:

Applied Language Learning
ATFL-AP-AJ
ATTN: Editor (Dr. L. Woytak)
Defense Language Institute
Foreign Language Center
Presidio of Monterey, CA 93944-5006

Manuscripts should be typed on one side only on 8-1/2 x 11 inch paper, double-spaced, with ample margins. Subheads should be used at reasonable intervals. Typescripts should typically run from 10 to 30 pages.

All material submitted for publication should conform to the style of the *Publication Manual of the American Psychological Association* (4th Ed., 1994) available from the American Psychological Association, P. O. Box 2710, Hyattsville, MD 20784.

Review Process

Manuscripts will be acknowledged by the editor upon receipt and subsequently sent to at least two reviewers whose area of expertise includes the subject of the manuscript. *Applied Language Learning* uses the blind review system. The names of reviewers will be published in the journal annually.

Specifications for Floppy Disks

Preferably use Windows-based software. Format manuscripts produced on one of the DOS-based or Macintosh systems, as an ASQII file at double density, if possible. Please name the software used. MS Word or text documents preferred.

Copyright

Further reproduction is not advisable. Whenever copyrighted materials are reproduced in this publication, copyright release has ordinarily been obtained for use in this specific issue. Requests for permission to reprint should be addressed to the Editor and should include author's permission.

Tear Out and Remove

From:

_____

_____

_____

Official Business

Place Stamp Here

Applied Language Learning
Defense Language Institute Foreign Language Center
Presidio of Monterey, CA 93944-5506
USA