The ACTFL Proficiency Guidelines: Gateway to Testing and Curriculum

Judith E. Liskin-Gasparro
Educational Testing Service

ABSTRACT The ACTFL Provisional Proficiency Guidelines, *a series of proficiency levels for speaking, listening, reading, writing, and culture in a foreign language, are a guiding principle for the development of communicative, proficiency-oriented tests and curricula. Based on proficiency descriptions originally developed by the U.S. government, the guidelines at each level include statements of characteristic linguistic functions, content areas, and accuracy. The features of proficiency and achievement tests are discussed, and two proficiency-based curriculum projects are described.*

Introduction

The *ACTFL Provisional Proficiency Guidelines,* developed with funding from the International Research and Studies Program of the U.S. Department of Education and published in 1982, consist of a series of descriptions of proficiency levels for speaking, listening, reading, writing, and culture in a foreign language. The set of guidelines introducing this issue includes language-general (generic) descriptions. The published guidelines also include language-specific guidelines for French, German, and Spanish. As of this writing, a second project is under way to develop similar proficiency descriptions for Chinese, Japanese, and Russian. The guidelines projects have been undertaken in

response to a recommendation of the President's Commission on Foreign Language and International Studies to establish "language proficiency achievement goals for the end of each year of study at all levels, with special attention to speaking proficiency."[1]

The ACTFL Guidelines were written as a first and essential step in the development of articulated end-of-course goals. As the introduction to the guidelines states, they address a bipartite need for proficiency-based curriculum development and evaluation of both students and programs:

> There is great potential for the impact of these guidelines on foreign language instruction. Measurable proficiency goals will form the basis for curriculum planning and classroom teaching. Students will more quickly develop a sense of accomplishment and will be able to refer to these "yardsticks" to measure their progress. The complex problem of articulation, the coordination of content (and skills) between grade and course levels, can also begin to be addressed. Student evaluation and placement can be based on actual language proficiency instead of on inaccurate and relatively uninformative measures of "seat time."[2]

The pages that follow describe the guidelines in greater detail and explore their application to two important areas of foreign language education—testing and curriculum development.

The Guidelines Revisited

As stated above, the guidelines are a sequential series of descriptions of the ability to speak, under-

Judith E. Liskin-Gasparro (ABD, Princeton University) is a Senior Examiner in the College Board Division at Educational Testing Service; Princeton, NJ.

stand, read, write, and operate culturally in a foreign language. They span a broad range of proficiency from "the most minimal acquaintance with the language to adult professional-level skill."[3] Many people who approach the guidelines for the first time find them somewhat cryptic and dense. The logic of an articulated sequence of objectives and their promise for improving instructional programs are apparent, but the structure of the guidelines and the composition of the level descriptions must be exposed and understood before one can begin to use them in an instructional setting.

History

A bit of history is in order here. Looking only at the immediate past, we might say that the guidelines are derived from the report of the President's Commission on Foreign Language and International Studies, and from the research and developmental efforts in functional-notional syllabi in which European foreign language educators have been involved for the last decade or so. Placing them in a larger context, it is possible to trace the guidelines as a conceptual construct through some 50 centuries of language teaching.[4]

It is probably most fruitful to look back some 30 years or so, to the beginning of systematic attempts by the Foreign Service Institute (FSI) of the U.S. Department of State to measure the outcomes of its language instructional programs. The process that was used to develop the descriptions of language ability has been described by several sources.[5] Interestingly enough, the procedures used to develop the oral proficiency interview test and rating scale some 30 years ago—careful job analysis, including direct observation of the skills required on the job; determination of what elements of the job (i.e. the spoken language) are crucial to success and the development of criteria for success on the job; and, finally, the design of a test that can measure whether individuals meet the established criteria—are virtually identical to the steps in the construction of vocational work-sample performance tests.[6] This striking similarity bespeaks the recognition and frank appreciation of foreign language proficiency as an essential skill in job performance.

The work undertaken at FSI in the 1950's established descriptions of language proficiency based on observable language skills required on the job. The original one-line definitions were eventually expanded to paragraph-length descriptions and were standardized and formally adopted in 1968 by the members of the Interagency Language Roundtable (ILR), a consortium of government agencies involved in language training and

evaluation.[7] In addition to the definition of levels of proficiency, the FSI staff, under the direction of Claudia Wilds,[8] devised a structured interview test to elicit from each examinee a representative sample of his or her speech.

In the 1960's, Educational Testing Service (ETS) personnel were trained by FSI to use the oral interview, and the State Department contracted with ETS to manage the training of oral proficiency testers for ACTION/Peace Corps. For close to two decades, ETS has conducted in-country tester training programs and has overseen the testing of thousands of volunteers in more than 60 languages.

In the last decade, the states of California, Florida, Illinois, New Jersey, and Texas have enacted an oral proficiency requirement for the certification for bilingual education teachers. ETS has conducted numerous tester training projects in those states and has assisted in decision making about proficiency standards, testing programs, and the like.

Until 1978 interest in the oral interview outside the government had been limited almost completely to the areas described above. In the period from 1978 to the present, there has been a virtual explosion of interest in academic circles in the concept of proficiency guidelines and proficiency-based curricula, materials, and evaluation. In addition to the President's Commission, the MLA-ACLS Language Task Force Projects, the ETS Common Metric Project, and the FSI Testing Kit Workshops all served to bring to the attention of the profession the need for proficiency-based guidelines.[9]
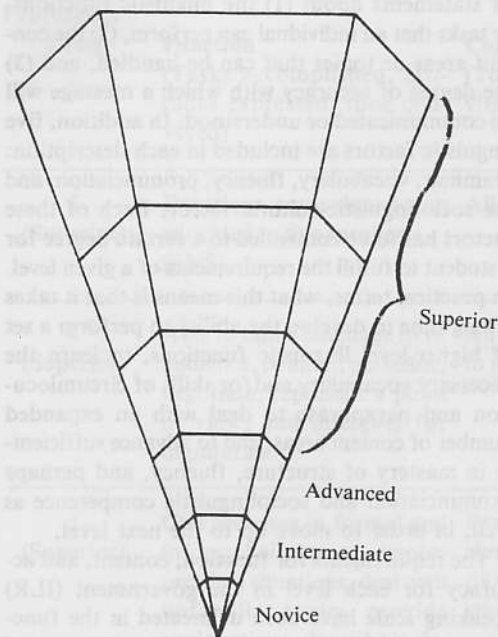
Explicating the Guidelines

To understand the ACTFL Guidelines more fully, it is helpful to return to the original one-line FSI speaking definitions adopted in 1956. The description of each level can be summarized as follows:

Level 0: no functional ability

Level 1: elementary (survival-level) proficiency

Level 2: limited working proficiency

Level 3: professional working proficiency

Level 4: full professional proficiency, or representational proficiency

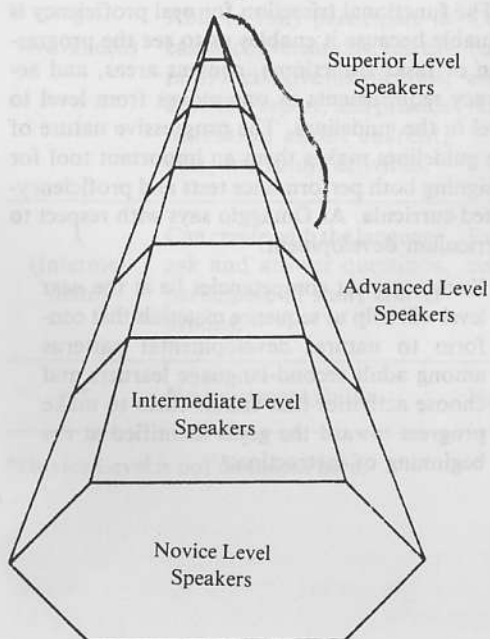Level 5: proficiency indistinguishable from that of an educated native speaker

The scale defined by these descriptions is not linear; it is best depicted as a cone that flairs dramatically outward as one ascends the scale. (See Figure 1.) This means that while it is relatively easy to move from 0 (no functional ability) to Level 1, the ability to survive linguistically by means of sim-

ple, tourist-type language, it becomes relatively more difficult and more time-consuming to move from level to level as one ascends the scale.

Figure 1: Inverted Pyramid of Language Proficiency Levels

Superior

Advanced

Intermediate

Novice

Figure 2: Relative Numbers of Speakers at Each Pyramid Level

Superior Level Speakers

Advanced Level Speakers

Intermediate Level Speakers

Novice Level Speakers

The cone that depicts the proficiency scale should be inverted to form a pyramid if we wish to think about the relative numbers of individuals at each level who claim some knowledge of another language. (See Figure 2.) At the bottom of the pyramid, where the broad base signifies a large number of language users at Levels 0 and 1, we find most students who are currently studying a language and the great majority of adults who studied a language in their high school and/or college days. Moving up the scale to Levels 2 and 3, where the pyramid narrows considerably, we find a much smaller number of individuals who have taken language study very seriously, and who have continued to use the language as an important part of their professional lives. In a study in 1967,[10] John Carroll administered a battery of tests, including the oral proficiency interview, to foreign language majors in their last semester of college. These students, who presumably had studied the language for 6-8 years, who may have had the additional advantage of study abroad, and who were probably expecting to use their language skills in their future professions, typically rated 2 or 2+.

At the very top of the scale we are approaching the point of the pyramid. Here there is an extremely small number of language users, those who are near-native or equivalent to educated native speakers. The people in this category have not acquired their language skills primarily through formal study, but rather through unusual biographical circumstances, such as speaking the language at home, residing for a long time in a country in which the language is spoken, engaging in advanced study through the medium of the language, and so on.

Since the great majority of students find themselves at the lower end of the language proficiency scale, ETS first, and then ACTFL and ETS working together, devised an academic version of the government scale. This scale is used by the ACTFL Guidelines; the workshops in oral proficiency testing conducted by ACTFL/ETS since 1982 have all been based on the academic scale as well.

While it articulates perfectly with the government scale and holds to the meaning of each level as defined by the ILR, the ACTFL/ETS scale has made three major innovations: (1) it has expanded the lower end by describing sub-ranges of proficiency within Levels 0 and 1; (2) it has correspondingly compressed the upper end of the scale by combining Levels 3-5 into one level denoting professional proficiency in the language; and (3) it has renamed the levels, so that Level 0 = Novice, Level 1 = Intermediate, Level 2 = Advanced, and Levels 3-5 = Superior.[11]

The reasons for these modifications lie in the nature of the proficiency scale itself and in the process of acquiring proficiency in a language. Whether we are talking about speaking, understanding, reading, or writing a foreign language, it takes a long time to move from one level of the scale to the next. Data from the FSI[12] indicate that it takes 240 hours of instruction for a professional adult with average language-learning aptitude to reach Level 1 in a language that is relatively close to English, such as French, Italian, or Spanish. To move to Level 2 requires 480 hours; even after 720 hours, proficiency no higher than Level 2+ can be expected.

As Liskin-Gasparro and Woodford[13] point out, academic language study takes place in less ideal conditions than those enjoyed by teachers and learners at government language schools. In government training programs, classes are quite small, only rarely numbering as many as ten students. Learning the language is the students' only activity; their 6-8 hours of daily instruction receive no competition from other courses, extracurricular school activities, or part-time jobs. The students may already be experienced language learners, working on their third or fourth foreign language. Most important, perhaps, is the motivation factor: government learners know that they will need the language after the course of study in order to carry out their jobs. In some agencies, salary increments are attached to attainment of a certain level of language proficiency.

In schools and colleges, on the other hand, students typically receive 70-100 hours of instruction per year. Many of them may not be in a foreign language class by choice, and certainly none of them are able to devote themselves to language study with the intense concentration of a foreign service officer. Even under the best of circumstances, it still takes over two years of instruction for students to accrue the 240 hours that bring their government counterparts to Level 1. And it may take the rest of their high school and college careers for them to reach Level 2.

Consequently, students spend a long time, in most cases all of their language-learning years, at Levels 0 and 1. The ACTFL/ETS scale provides additional proficiency benchmarks so that progress within these two levels can be measured. The ranges are themselves so broad that under the government scale students might improve considerably over the course of a semester or a year and yet not have that improvement reflected in a higher rating.

The Functional Trisection

One might well ask why the ranges are so broad, and why so much learning must take place in order to move from one level to the next. The principal reason is that each level is composed of a complex constellation of factors. An analysis of the level descriptions reveals that each of them is composed of statements about (1) the linguistic functions, or tasks that an individual can perform, (2) the content areas or topics that can be handled, and (3) the degree of accuracy with which a message will be communicated or understood. In addition, five linguistic factors are included in each description: grammar, vocabulary, fluency, pronunciation, and the sociolinguistic-cultural factor. Each of these factors has to be controlled to a certain degree for a student to fulfill the requirements of a given level. In practical terms, what this means is that it takes a long time to develop the ability to perform a set of higher-level linguistic functions, to learn the necessary vocabulary and/or skills of circumlocution and paraphrase to deal with an expanded number of content areas, and to advance sufficiently in mastery of structure, fluency, and perhaps pronunciation and sociolinguistic competence as well, in order to move up to the next level.

The requirements for function, content, and accuracy for each level in the government (ILR) speaking scale have been delineated in the functional trisection.[14] (See Figure 3.) Development of similar outlines for the skills and factors involved in performance at each level for the other skills is under way.

The functional trisection for oral proficiency is valuable because it enables us to see the progression of tasks (functions), content areas, and accuracy requirements as one moves from level to level in the guidelines. The progressive nature of the guidelines makes them an important tool for designing both performance tests and proficiency-based curricula. As Omaggio says with respect to curriculum development:

> Knowing what competencies lie at the *next* level will help us sequence materials that conform to natural developmental patterns among adult second-language learners and choose activities that enable them to make progress toward the goals identified at the beginning of instruction.[15]

Figure 3
FUNCTIONAL TRISECTION OF ORAL PROFICIENCY LEVELS

Oral Proficiency Level	Function (Tasks accomplished, attitudes expressed, tone conveyed)	Context (Topics, subject areas, activities and jobs addressed)	Accuracy (Acceptability, quality, and accuracy of message conveyed)
5 (Superior)	Functions equivalent to an educated native speaker (ENS).	All subjects.	Performance equivalent to an ENS.
4 (Superior)	Able to tailor language to fit audience, counsel, persuade, negotiate, represent a point of view, and interpret for dignitaries.	All topics normally pertinent to professional needs.	Nearly equivalent to an ENS. Speech is extensive, precise, appropriate to every occasion with only occasional errors.
3 (Superior)	Can converse in formal and informal situations, resolve problem situations, deal with unfamiliar topics, provide explanations, describe in detail, offer supported opinions, and hypothesize.	Practical, social, professional, and abstract topics, particular interests, and special fields of competence.	Errors never interfere with understanding and rarely disturb the native speaker. Only sporadic errors in basic structures.
2 (Advanced)	Able to fully participate in casual conversations, can express facts, give instructions, describe, report, and provide narration about current, past, and future activities.	Concrete topics such as own background, family, interests, work, travel, and current events.	Understandable to native speaker *not* used to dealing with foreigners, sometimes miscommunicates.
1 (Intermediate)	Can create with the language, ask and answer questions, participate in short conversations.	Everyday survival topics and courtesy requirements.	Intelligible to native speaker used to dealing with foreigners.
0*	No functional ability.	None.	Unintelligible.

*Novice Level is not discussed here.

Function, Content, and Accuracy
at the Intermediate Level

Let us take the Intermediate Level for speaking proficiency as an example. It is an important level, since most students emerging from four years of secondary school or four semesters of college study will fall somewhere in this range. The original FSI designation for Level 1, or the Intermediate Level, is "elementary (survival) proficiency." This means that an individual who speaks a language even at the low end of the Intermediate Level will be able to survive linguistically for a day or two in an environment in which the language is spoken in the way a tourist might speak it. The individual will be able to take care of such basic needs as ordering a meal, getting lodging in a hotel, mailing a package, making a simple purchase, and getting directions.

The level description for Intermediate Low on the ACTFL/ETS scale reads as follows:

Able to satisfy basic survival needs and minimum courtesy requirements. In areas of immediate need or very familiar topics, can ask and answer simple questions, initiate and respond to simple statements, and maintain very simple face-to-face conversations. When asked to do so, is able to formulate some questions with limited constructions and much inaccuracy. Almost every utterance contains fractured syntax and other grammatical errors. Vocabulary inadequate to express anything but the most elementary needs. Strong interference from native language occurs in articulation, stress, and intonation. Misunderstandings frequently arise from limited vocabulary and grammar and erroneous phonology, but, with repetition, can generally be understood by native speakers in regular contact with foreigners attempting to speak their language. Little precision in information owing to tentative state of grammatical development and little or no use of modifiers.

If we analyze this description, we can see that every statement in it refers to the function, context/content, or accuracy of speakers at this level. The functional trisection for the Intermediate Level reads as follows:

Function: Can create with the language; ask and answer questions, participate in short conversations.
Content: Everyday survival topics and courtesy requirements.
Accuracy: Intelligible to a native speaker used to dealing with foreigners.

In the area of function, the level description above contains the following statements:

1.   Able to satisfy basic survival needs and minimum courtesy requirements. (This is a summary statement.)
2.   Can ask and answer simple questions, initiate and respond to simple statements, and maintain very simple face-to-face conversations. (This is almost identical to the language of the functional trisection.)

All of the function statements can be summarized as the ability to *create with the language.* Speakers at this level are able to manipulate the bits and pieces of language under their control to form sentences to communicate their own meaning. Speakers at the Novice Level are not able to do this; they are limited to the elements they have memorized, and can only deliver this memorized material if the context presented to them is identical or very similar to the context in which the utterances were learned. For example, a Novice Level "conversation" on the subject of the family might go something like this:

Teacher: Tell me about your family.
Student: (Silence)
Teacher: How many people are in your family?
Student: Four.
Teacher: Who are they?
Student: Mother. Father. Brother. Me.
Teacher: Tell me something about your brother.
Student: (Silence)
Teacher: How old is your brother?
Student: 14.
Teacher: What is his name?
Student: John.
etc., etc.

At the Intermediate Level, on the other hand, students have sufficient proficiency to join linguistic elements together into simple sentences in order to communicate information about themselves. The same conversation at the Intermediate Level might sound like this:

Teacher: Tell me about your family.
Student: I have four people in my family. I have a mother. I have a father. I have a brother.
Teacher;   Tell me about your brother.
Student: My brother's name is John. He is 14. He plays football. He plays the violin.
etc., etc.

The student's responses may be just a series of short, discrete, structurally parallel statements. They may contain numerous agreement errors, syntax problems, and other inaccuracies, but the critical factor, from a functional point of view, is that the student has crossed an all-important

threshold from operating with memorized material and isolated words and phrases at the Novice Level to creating with language at the Intermediate Level.

In the area of content, the description of the Intermediate Level contains the following statement:

In areas of immediate need or very familiar topics...

What constitutes "familiar topics" will be determined by students' language-learning experience, but by and large students at the Intermediate Level can talk about such things as their families, pets, homes, friends, school schedules, interests, and likes and dislikes.

The rest of the Intermediate Level description concerns accuracy:

1. ...is able to formulate some questions with limited constructions and much inaccuracy.
2. Almost every utterance contains fractured syntax and other grammatical errors.
3. Vocabulary inadequate to express anything but the most elementary needs.
4. Strong interference from native language occurs in articulation...
5. Misunderstandings frequently arise...but, with repetition, can generally be understood by native speakers in regular contact with foreigners...
6. Little precision in information.

All of these statements can be summarized by one line in the functional trisection: "intelligible to native speaker used to dealing with foreigners." Accuracy includes structure, vocabulary, pronunciation, and even fluency. Note that at this level a sympathetic, linguistically knowledgeable interlocutor is assumed.

The Guidelines as Definitions of Proficiency

An interesting aspect of the guidelines, and one that is often misunderstood, is that they are what the FSI terms "absolute definitions" of proficiency. Factors such as length of study and the relative difficulty of the language for speakers of English are irrelevant. An Intermediate Low speaker of Japanese has the same level of proficiency as an Intermediate Low speaker of German; both will be able to carry out the same functions, talk about the same kinds of subjects, and will exhibit the same degree of accuracy with respect to the intelligibility of their speech to a native speaker used to dealing with foreigners. The difference is that it will probably take an American student of Japanese

longer to reach that level, because Japanese is more distant from English than is German.

Within the Novice and Intermediate Levels, the levels for which ACTFL/ETS has elaborated additional proficiency descriptions, speakers continue to fit within the parameters defined by the functional trisection. The difference between Novice Low and Novice Mid speakers is purely one of quantity; speakers at Novice Mid will have more isolated words and phrases than speakers at Novice Low. At the Intermediate Level, the difference between Low and Mid performances is one of both quantity and quality. Speakers at Intermediate Mid can say many things simply, and they are generally accurate in very basic constructions, such as subject-verb agreement for the present tense of regular verbs. Novice High, Intermediate High, and Advanced+ are equivalent to the ILR scale's 0+, 1+, and 2+, respectively. They indicate performance that has substantial features of the next level, but performance at that next level is uneven, not sustained.

Although the proficiency descriptions for speaking proficiency are the most completely developed and the best understood, and although much of the activity in the profession has focused on the assessment of students' oral proficiency in accord with the guidelines, it is important to keep in mind that proficiency guidelines exist for the other skills as well, and that these guidelines can be used to structure testing and instructional activities.

Evaluating Proficiency with the Guidelines

The chief way that the guidelines have been used in academia since their publication in 1982 has been to evaluate students' language proficiency. Although workshops and training programs have focused primarily on the evaluation of students' speaking ability, some first forays have also been made into the relatively virgin territories of listening, reading, and writing proficiency. The most significant of these to date was an invitational symposium on the receptive skills, sponsored by ACTFL and the National Security Agency, that was held in November, 1983. Participants were asked either to present or critique papers on five topics: "Similarities and Differences in the Receptive Skills," "Reading Proficiency," "Listening Comprehension Proficiency," "Testing Proficiency in the Receptive Skills," and "Computer-Assisted Teaching and Testing."[16] Future plans include the development of computer-adaptive tests of listening and reading proficiency.

In a more modest undertaking, Liskin-Gasparro and Lowe made a day-long workshop presentation

on reading and writing proficiency to the Tennessee Foreign Language Teaching Association at its annual meeting in November, 1983. In a series of plenary and language-specific group sessions, Tennessee teachers became familiar with the ACTFL Proficiency Guidelines for reading and writing, and assigned levels to printed texts and student compositions in French, German, Spanish, and English. In the sessions on reading, various aspects of testing were discussed: (1) how to rate a text that will be used as the basis for a reading comprehension test; (2) how to write questions keyed to various levels of proficiency; and (3) how to evaluate students' responses. For the writing sessions, teachers had been asked to bring samples of student compositions. These were thrown into a large box in the registration area. The two workshop leaders sorted through the papers and selected several at each level on the scale, and then transferred them to transparencies for use in the next day's sessions. In the discussion, it became clear that the guidelines could be extremely valuable in devising essay topics and in comparing writing samples from various students. Remarkable consensus was achieved in the rating of the compositions in sessions only an hour in length.

The value of proficiency tests is that they measure, by definition, real-life language ability. They imitate the language tasks that an individual would have to carry out in life outside the classroom—while traveling abroad, working in a business with international connections, working in a bilingual setting in the United States, and so on. Since they measure language ability in relatively broad terms, proficiency tests should not be administered very often. They are most appropriately used to measure students' ability at crucial points such as the end of high school or college study, or after a significant language-learning experience, such as a semester or a year abroad.

Proficiency tests are being used in several language departments at the University of Pennsylvania as a means of assessing whether students have fulfilled the language requirement.[17] Penn is in the process of converting to a proficiency-based requirement from the traditional four-semester, seat-time requirement. This new system requires that students actually attain a certain level of proficiency in a language in order to fulfill the requirement; passing language courses is no longer sufficient. Students' proficiency is assessed by a series of proficiency examinations in the four skill modalities and culture.

Levels of Proficiency

If an institution or an agency is considering the use of proficiency testing for the language requirement or for certification for teachers, the question of standards is of crucial importance. How proficient should one be to complete a language requirement, teach in a bilingual or foreign language classroom, or work as a bilingual secretary? One of the most misleading phrases to come out of the recent focus on language proficiency is to say that an individual is *proficient* in a language. The ACTFL/ETS scale provides for descriptions of a wide variety of usable proficiencies, from that needed to buy a bus ticket to Milan to that needed to comprehend a passage on Hegel's dialectic in the original German. The scales, and the testing procedures that accompany them, provide us with a full picture of an individual's proficiency and allow us to state with some confidence the linguistic functions the individual can carry out, the content areas that she or he can communicate about, and with what degree of accuracy. Higgs and Clifford make this point when, after reviewing the literature on communicative competence, they state:

> The question that needs to be asked is not merely "Was the student able to communicate?" but "*What* was [the student] able to communicate, and *how well?*" The *what* requires consideration of both the topic or context of the communication and of the language function that must be performed in that context. The *how well* entails judgment of linguistic accuracy and cultural authenticity...A student cannot merely be declared competent in communication. The function that [the student] is competent to express must be specified...[18]

Proficiency Tests and Achievement Tests

To measure students' progress in a course, achievement tests are much more appropriate than proficiency tests. While proficiency tests measure real-life competence, achievement tests are anchored to a particular curriculum. Their purpose is to determine how well students have learned a particular body of material that they have been taught. The best achievement tests are those whose format and content correspond to the activities used and the topics covered in the course itself. For example, if students' exposure to the spoken language has consisted solely of listening to the teacher read graded materials from the text, it would not be good achievement testing procedure to include on the final exam comprehension questions about dialogues recorded by unfamiliar voices. A proficiency test, on the other hand, is not designed to find out what students have learned in a course,

but rather to find out how well the students' language will serve them in real-life situations. Clark expresses it this way:

"...in proficiency testing, the frame of reference—both for test development and score interpretation—shifts from the classroom to the actual situation in which the language is used. The student's performance on the test is analyzed not in terms of "how much [the student] has learned" but in terms of how clearly [the student's] performance meets the goal standards specified.[19]

It is clear that achievement testing is a vital part of all good language programs. The teacher needs to know at fairly regular intervals if the students are learning the material that is being taught. A well-designed achievement test serves diagnostic purposes as well, and allows the teacher to pick out those areas in which students need additional instruction and practice.

Proficiency testing serves an equally valuable role in a language program, that of providing an outside perspective and a check that the instructional goals, methods, and outcomes are all synchronized. The relationship between what students are learning and how well it will actually serve them in the world outside the classroom is fuzzy at best. Traditional classroom practice, such as pattern drills, discrete sentence fill-ins, and guided compositions give an illusory sense of performance that is most often not borne out when students are put in a position of having to communicate in the foreign language without the assistance of teacher, textbook, and highly structured exercises. In addition, proficiency testing emphasizes the viewpoint that languages are learned for a purpose—to be used. A proficiency test allows us to see how students are progressing along that path.

Designing a Proficiency Test

This section describes some characteristic features of proficiency tests, and might even be of use to teachers making up classroom tests—even though these, as noted above, are most often achievement tests, not proficiency tests. The guiding principles discussed here can assist teachers in making their tests more thematically unified and, if they so desire, more closely aligned with the ACTFL Guidelines. The goal of all test development is, of course, to create an instrument that responds to the teacher's assessment needs.

The traditional triad of characteristics that all good tests should have are validity, reliability, and practicality.

Validity refers to a test's ability to measure what it purports to measure. The oral interview, for example, is a highly valid test of speaking ability. Performance on the test and the ability to speak at the level of one's rating in the real world have been examined, compared, and agreed upon in over 30 years of government use. In addition, the oral interview enjoys what is known as *face validity*; that is, the examinee and the test user have a sense of confidence in the test as a valid measure of speaking ability.

Good tests must also possess a high degree of *reliability,* i.e. consistency of results in a test-retest situation. When it comes to reliability, the machine-scorable discrete-point test is king. If the test items are well-constructed (unambiguous, fair, etc.), then the reliability will be very high, since there will be no inconsistency in the scoring. A performance test, such as an oral interview or a series of writing samples, has to be scored by relatively unreliable human beings, and so will yield less consistent results than will machine-scorable tests. Even so, experience with the oral interview and with the holistic scoring of student compositions by English teachers indicates that thoroughly trained evaluators can reach high levels of intra-rater and inter-rater reliability.[20]

A third feature of good tests is *practicality,* which involves such factors as cost, time, and ease of scoring. Direct tests of writing and speaking ability will never be truly "practical," in the sense that they will always require students to produce language, which in turn will have to be evaluated by human raters. On the other hand, proficiency tests of receptive skills can be developed in machine-scorable format. As of this writing, the Defense Language Institute and the Language Schools of the Central Intelligence Agency and the National Security Agency all use multiple-choice reading proficiency tests that are extremely practical to administer and score. They have been validated against direct tests of reading proficiency, so that it is possible to convert a multiple-choice test score to a proficiency rating.

Proficiency test development follows another set of guiding principles in addition to the more general ones described above. Canale[21] has described four general test design features that apply to the development of proficiency tests: (1) thematic organization; (2) four stages in test administration; (3) adaptive testing procedures; and (4) criterion-referenced testing.

Canale advocates here that tests be designed thematically in order to "provide a coherent, natural, and motivating structure to the overall

test.'' This notion is in direct contrast to practices in most standardized and even teacher-made tests, which are organized according to linguistic criteria (vocabulary, structure, etc. or even, at the classroom level, by various grammatical topics) and which present students with a series of thematically unrelated sentences as stimuli. Students are, unfortunately, trained to cope with and even expect this shotgun approach by their textbooks, most of which have traditionally organized exercises and activities as though the target language were a series of non sequiturs.[22]

A well-constructed oral interview is thematically connected in the way Canale suggests. The interviewer will weave questions around a limited number of topics of interest to the examinee. The greater the skill of the interviewer, the more thematically connected the interview will be, while still covering a wide range of topics and functions.

Canale's second recommendation is that proficiency tests use a four-stage administration approach. This is a suggestion drawn from the oral interview, which has four well-defined phases, each with its own set of purposes. The first phase, or warm-up, consists of questions that should be quite easy for the student, and are intended to put him or her at ease. The next phase, the level check, is the tester's attempt to find the level of language at which the student sustains solid, comfortable performance. Having found this level, the tester moves one level higher to the probes, in order to present the student with tasks at a level just beyond his or her level of proficiency. The purpose of the probes is to verify the maximum proficiency level by indicating the ceiling, that is, the point beyond which the student cannot go. This serves to verify the rating both to the tester and to the student, who will be aware of his or her limitations. The wind-down returns the student to a comfortable performance level so that the interview ends on a positive note.

The four-stage approach can be applied to proficiency tests of the other skills as well. They should begin with easy questions, work up the scale to the highest level that the student can handle comfortably and competently, and then move beyond that level to check that it is indeed the highest level of performance that the student can sustain. They will end with questions at the student's proficiency level, in order to confirm the rating one last time and to give the student a sense of accomplishment.

Canale's third recommendation is that proficiency measures use adaptive test procedures. The oral interview is the adaptive test *par excellence*. The tester begins with an overall test structure, com-
mon to all interviews; but within that structure the test is tailor-made for each examinee. In the area of content, no two tests are exactly alike, since the tester will seek information on the student's functional ability using questions based on topics inspired by the student's responses in the course of the interview. The oral interview is also adaptive in a vertical sense; the interviewer adjusts the level of the test tasks demanded of the student according to the proficiency level demonstrated by the student in his or her previous responses.

Plans are under way at ACTFL and ETS to develop computerized adaptive tests of listening and reading proficiency. The computer would serve as the technological analogue of the human tester in an oral interview. Just as the tester rates the student's performance throughout the interview, adapting the content of questions, the question types, and the difficulty level according to the student's responses, the computer would be programmed to do the same kind of evaluation and the same kind of test task and item selection.

Test developers and researchers at ETS and elsewhere who have been actively involved in the theory and technical considerations have pointed out numerous benefits of computerized adaptive tests over the traditional test booklet/answer sheet variety:[23]

1) Adaptive tests are coherently more interesting to the student, since the questions are focused at each student's level of ability. In fact, the purpose of the adaptive program is for the computer to quickly zero in on the student's precise level of ability. In traditional tests, which have questions spanning a wide range of difficulty, students will be bored by questions that are too easy for them and frustrated by those that are far too difficult.

2) Since adaptive tests only present students with questions at their ability level, the tests can be far shorter than tests made to serve a heterogeneous student population.

3) Since the test is administered by the computer, there is no need for answer sheets which, as we all know, can easily be torn, spindled, mutilated, or lost. In addition, students can work at their own pace. The tests can be scored immediately, allowing them to be used diagnostically.

4) Test security is improved, since test questions can be stored more safely in a computer than in a locked office.

5) Test development functions can be handled more easily. Pretested items can be inserted and tried out, and items found to be defective can quickly be removed without massive and expensive

reprinting.

6) As the technology improves, the possibility for test formats other than multiple-choice will increase dramatically. Students can already type in their answers; voice synthesizers and/or synchronized videotapes will allow for the testing of comprehension of various types of oral language.

Finally, Canale states that proficiency tests should be criterion-referenced. Although the proficiency guidelines and tests built to measure proficiency are by definition criterion-referenced, it is important to reemphasize this essential feature. The classic definition of a criterion-referenced test is a test designed to measure to what extent a student can perform a particular task. Criteria are established, test questions are constructed to reflect the criteria, and students' performance is evaluated *with respect to the criteria.*

Criterion-referenced tests are a relatively recent arrival on the educational scene, and are usually contrasted with the older and more well-established norm-referenced tests. The purpose of a norm-referenced test is to rank groups of students. The questions are constructed over a wide range of difficulty so that students will be spread across the whole score scale according to their performance *with respect to each other.*

The value of a criterion-referenced test is twofold:

1) in order to design a criterion-referenced test, it is necessary to begin with a well-defined set of objectives; and

2) a student's performance on a criterion-referenced test is interpreted without regard for the performance of other students.

It is often the case, however, that a test can become norm-referenced or criterion-referenced according to the interpretation of test results. A proficiency test rated on the Novice-Superior scale is less likely than most other tests to be used to rank students because there are so few ratings on the score scale. If an instructor were to misinterpret the intention of the oral proficiency interview and assign grades in a course on the basis of performance (Intermediate High = A, Intermediate Mid = B+, Intermediate Low = B, Novice High = C+, etc.), then that instructor would be erroneously making the oral interview into a norm-referenced test, in spite of its criterion-referenced nature.

Using the Guidelines
for Curriculum Development

It has long been recognized that tests are a driving force in curriculum. If students are to take a particular type of standardized test at the end of a course or program of studies, teachers will understandably begin to "teach to the test," to modify their curricula so that students will learn what they need to know to perform well on the test. When the test in question is unrelated to the goals and methods that the teacher believes in, then "teaching to the test" can be an educational disaster. If, on the other hand, students take proficiency tests at the end of their sequence of language courses, tests that will assess how well they can handle the language in real-life situations, then the pressure to "teach to the test" can assist the teacher in developing a more proficiency-oriented program.

The pioneering work of using the ACTFL Guidelines as an organizing principle for curriculum development began in the summer of 1983, when ACTFL held an NEH-sponsored three-week summer institute for secondary school teachers on proficiency-based curriculum development. In the first week, the participants were trained to conduct and rate oral proficiency interviews. Weeks 2 and 3 were spent with their textbooks, their curriculum guides, and the ACTFL Guidelines, as they developed projects to revise their own curricula or to make classroom activities more communicative.

The teachers organized themselves into small groups, and began work on projects that were to be tried out and refined during the fall semester. Most of the projects involved classroom activities, and were focused on moving students over a critical threshold on the proficiency scale. Several Spanish teachers, who came to be known as the "preterite group," had noticed in the course of the oral interviews in Week 1 that although students learn to form the preterite in Spanish at the end of the first year, and presumably review the use of the preterite in second year and thereafter, even advanced students in fourth- and fifth-year courses are unable to use the preterite naturally in narrations and descriptions in past time. (This is an essential function at the Advanced Level.) Consequently, the "preterite group" devised a series of 16 activities involving the use of the preterite in communicative situations. The activities were to be used at intervals in second-, third-, and fourth-year courses in an attempt to give students additional practice with narration in past time.

One group of four French teachers who use Valette and Valette's *French for Mastery*[24] focused on the Novice and Intermediate Levels, and devised a series of structured role plays and small-group activities around the material in the text that would allow even first-year students to use the

language orally. Teachers' attempts at conversation or even simple role plays in first-year classes are usually met with silence or giggles rather than language because students are simply not yet able to create with language; they can only say what they have learned in the context in which it was presented to them. The role plays follow the content of the text quite closely, so that students who have learned the material can communicate easily; the value lies in practicing conversational turns and in using the language orally. The small-group activities, in which groups of 2-4 students are actively engaged in exchanging information and recording their partners' responses for subsequent reports to the class, increase dramatically the amount of language generated in a single class period. Some of the activities are structured so that students get to practice functions of the Intermediate Level, such as asking questions and giving information on familiar topics. The tight structuring of the exercises makes it possible for even Novice Level speakers to perform them.

A second class of projects involved using the proficiency guidelines as starting points for the development of curricula. In this regard, it is important to remember that the ACTFL Guidelines are not curriculum outlines, nor are they prescriptions for what grammatical structures to teach and when. They are a graduated sequence of proficiency stages around which a foreign language program may be structured. The day-to-day activities that constitute the sequence of small steps in the context of the larger phases identified by the guidelines, as well as the methods to be used, are still and always will be the province of the foreign language teacher.

Perhaps the most ambitious curriculum-development undertaking to emerge from the ACTFL summer institute is a new set of curriculum guidelines for the French program at Walpole (MA) High School, designed by teachers Floy Miller and Charlotte Cole. In the introduction to the first version of their curriculum outlines for French I, French II, and French II Honors,[25] Miller and Cole

Figure 4

CURRICULUM GUIDELINES: LEVEL I

	Function	Content
SPEAKING	Able to express basic courtesy formulae. Able to make short statements. Able to enumerate in short phrases.	Immediate needs such as greetings, basic objects, days, months, colors, weather, etc.
LISTENING	Able to comprehend basic courtesy formulae. Able to comprehend basic vocabulary and memorized material. Able to comprehend utterances in highly contextualized situations.	Immediate needs such as greetings, basic objects, months, days, colors, weather, etc. Immediate needs such as in a restaurant, store, train, bus, etc.
READING	Able to read basic vocabulary and short phrases. Able to read instructional and directional material. Able to read mastered material or recombinations of mastered material.	Immediate needs such as names, addresses, signs indicating names of streets and avenues, building names, and short informative signs. Immediate needs such as items on menus, schedules, timetables, maps, and social codes (*Défense de fumer,* etc.). Basic survival needs such as messages, greetings, and social amenities in dialogues and specially prepared texts.
WRITING	Able to copy isolated words and short phrases. Able to transcribe memorized words and phrases. Able to write memorized words, lists, and phrases.	Immediate needs such as name, address, dates, numbers, days, months, and common objects. Courtesy formulae such as expressions in greetings and leave-takings.

explain that they wished to escape from the futile exercise that curriculum writing so often represents: "...we did not wish to produce curriculums that had no value for the classroom or ones that had little use other than cosmetic for the shelves or files where such documents seem inevitably to rest in peace until the next revision."

Miller and Cole began by examining the departmental goals for the language program as a whole, and interpreting them in light of the ACTFL Guidelines. They then devised end-of-year goals for speaking, listening, reading, writing, and culture for all five course levels in French. The end-of-course goals are also expressed in terms of the guidelines, following a sequence that will bring students who complete French V to a level of proficiency consonant with the overall program goals set by the department. The next step was to analyze the text used in each course to see how the material in each unit contributed to the students' attainment of the end-of-course proficiency goals. Each unit was outlined, and day-to-day communicative "implementation" activities were designed to make the textbook serve more fully the goals of the course.

For illustration, the proposed French I curriculum guidelines are included. (See Figure 4.) The function statements are taken from the ACTFL Guidelines (note that students completing a first-year course are expected to be at the Novice Level in speaking, listening, and writing, and at the Intermediate Level in reading); the content statements are inspired by the Guidelines, but the details of content come from the topics and vocabulary covered in the textbook.

A perusal of the unit outlines for French I reveals that students are exposed to material beyond that which they are expected to master by the end of the course. For example, no ability to handle the past tense is expected by the end of French I, yet the *passé composé, imparfait,* and *futur* are all introduced during the course. The instructional approach is cyclical; students will be introduced to more advanced structures and will practice manipulating them mechanically in structured drill activities; at a later stage, they will be exposed to readings containing these structures, and will be expected to use them in structured writing exercises, such as guided compositions. At an even later stage, they may be expected to use them easily and naturally in conversation and in writing. By this time their familiarity with the structures and their general level of proficiency will have increased, making it possible for them to do so. This long-term, cyclical approach to the building of proficiency is supported by Omaggio, who writes:

[In a proficiency-oriented approach] contextualized and meaningful drills...allow students to practice specific structures or vocabulary in a way that has some psychological reality and communicative value, and should therefore be used as soon as possible in the early levels of instruction...Open-ended, creative, and personalized practice...should follow this structured practice as soon as possible in a proficiency-oriented approach.[26]

Miller and Cole plan to use the oral proficiency interview primarily to evaluate their program, rather than to evaluate the proficiency of all of their students. The chief objection to the use of the oral interview with all language students is that it is impractical. No teacher has enough time to administer interviews to all of his or her students. Miller and Cole plan to assess students' oral proficiency as a tool to monitor their curriculum development project in the following ways:

(1) interview a stratified sample of students at the beginning of French II and French II Honors to determine the proficiency levels of students who are beginning language study at the secondary level;

(2) administer the oral interview to transfer students as an aid in placement;

(3) interview all seniors and underclassmen who are terminating language study, in the hope eventually of recording the rating on the students' transcripts; and

(4) interview a random sample of students who have completed a two-year sequence, in order to monitor progress and assess the effectiveness of the program.

The major innovation of Miller and Cole's work from a curriculum development standpoint is that their point of departure was the ACTFL Guidelines, an external set of statements of proficiency. They then modified the textbook to suit their needs, creating their own materials as necessary to fill in the gaps. Most curriculum guidelines take as their point of departure the table of contents of the textbook, for two main reasons: (1) until the publication of the ACTFL Guidelines, there were no systematically articulated goals to follow other than those implied by the sequence of chapters and books in the publishers' text series, and (2) most teachers have not been able to devote the kind of time to the mammoth endeavor of restructuring a program that Miller and Cole have.

The work undertaken by Miller and Cole at the secondary level has been paralleled by Isabelle

Kaplan of the French Department at Northwestern University. Her curriculum project, described elsewhere in this issue, was the creation of a conversation course based on the ACTFL Guidelines and designed to move her students from the Intermediate Level to the Advanced Level on the speaking scale.

Conclusion

In the final analysis, how can the ACTFL Guidelines and the proficiency tests based on them work together to affect curriculum? There are several ways.

First, and most important, projects like Miller and Cole's and Kaplan's utilize the guidelines directly as the organizing principle to design a curriculum. Proficiency tests in both of these projects were used not to evaluate students' performance for purposes of assigning grades, but rather to generate data on students' level of proficiency as an aid in the curriculum development work.

Second, tests in and of themselves drive instruction. This is a relatively recent phenomenon, which Popham traces back to 1965 and the passage of the Elementary and Secondary Education Act. According to Popham, the law required that funded programs be evaluated for quality and effectiveness in order to ensure continued federal support. This need for evaluation of students, not in comparison with one another but rather with respect to the program goals specified in the proposals for funding, sparked the proliferation of criterion-referenced tests in the decade that followed.[27] Over the last 20 years, external test scores have enjoyed an importance previously unknown in educational circles. The result has been that consciously or unconsciously, teachers orient their courses to the tests that will follow.

From the students' point of view, testing can have a positive backwash effect. If the students know, for example, that they will have periodic oral tests, they will begin to take oral production more seriously.

Third, foreign language courses organized around functional goals are more effective. Students' efforts to learn a particular structure will not be wasted if the material is introduced at a point and in such a way that students are able to incorporate it into their growing base of knowledge about and skill in the language.

The most basic, and perhaps most important goal that is inspiring the effort to promote proficiency-based foreign language education is that students emerge from language study with a usable level of language skill. The minimum level of skill that can be characterized as "usable" in this context is that which is needed to read a simple paragraph, carry on simple social conversations, and get from point A to point B in a country in which the language is spoken. Students who achieve this level of basic proficiency may well be motivated to learn more.

The role of the ACTFL Guidelines in this endeavor is to serve as an organizing principle for the development of proficiency-based curricula. With the consensus that is growing within the profession on the value of guidelines, teachers will be able to combine their own philosophies, methods, and experience with the curriculum structure suggested by the guidelines to build even more effective language programs.

NOTES

[1] *Strength Through Wisdom: A Critique of U.S. Capability. A Report to the President from the President's Commission on Foreign Language and International Studies* (Washington, DC: U.S. Government Printing Office, 1979).

[2] *ACTFL Provisional Proficiency Guidelines* (Hastings-on-Hudson, NY: American Council on the Teaching of Foreign Languages, 1982), p. 1.

[3] Judith E. Liskin-Gasparro, "The ACTFL Proficiency Guidelines: An Historical Perspective," in *Teaching for Proficiency, the Organizing Principle,* ed. Theodore V. Higgs, ACTFL Foreign Language Education Series, No. 15 (Lincolnwood, IL: National Textbook Co., 1984), p. 12.

[4] See Liskin-Gasparro, pp. 13-31.

[5] See Howard E. Sollenberger, "Development and Current Use of the FSI Oral Interview Test," in *Direct Testing of Speaking Proficiency: Theory and Application,* ed. John L.D. Clark (Princeton, NJ: Educational Testing Service, 1978), pp. 1-12; Liskin-Gasparro, pp. 18-19; and Pardee Lowe, Jr., "The ILR Oral Interview: Origins, Applications, Pitfalls, and Implications," *Die Unterrichtspraxis* 16 (1983), 230-32.

[6] Joseph L. Boyd, Jr. and Benjamin Shimberg, *Handbook of Performance Testing. A Practical Guide for Test Makers* (Princeton, NJ: Educational Testing Service, 1971), pp. 1-29.

[7] Work on the government descriptions did not end at that point, but rather has continued up to the present. The Testing Committee of the ILR, formally charged with the responsibility of monitoring the definitions and revising them as appropriate, completed early this year the newest set of descriptions of proficiency levels for speaking, listening, reading, and writing. This most recent version includes descriptions of the plus levels and provides illustrative examples of performance at each level.

[8] See Claudia Wilds, "The Oral Interview Test," in *Testing Language Proficiency,* ed. Randall L. Jones and Bernard Spolsky (Arlington, VA: Center for Applied Linguistics, 1975), pp. 29-38.

[9] For a description of these projects, see Liskin-Gasparro, pp. 23-31.

[10] John B. Carroll, "The Foreign Language Proficiency Levels Attained by Language Majors Near Graduation from College," *Foreign Language Annals,* 1 (1967), 131-51.

[11] For further discussion, see Liskin-Gasparro, pp. 27-28 and Lowe, pp. 232-33.

[12] "Expected Levels of Absolute Proficiency in Languages Taught at the Foreign Service Institute" (mimeo), Roslyn, VA: Foreign Service Institute, 1973.

[13] Judith E. Liskin-Gasparro and Protase E. Woodford, "Proficiency Testing in Second Language Classrooms," in *The Foreign Language Teacher: The Lifelong Learner,* ed. Robert G. Mead, Jr., (Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages, 1982), pp. 130-31.

[14] The source is Pardee Lowe, Jr., *Manual for LS Oral Interview Workshops* (Washington, DC: CIA Language School, 1980). Reprinted with permission.

[15] Alice C. Omaggio, "The Proficiency-Oriented Classroom," in *Teaching for Proficiency, the Organizing Principle,* ed. Theodore V. Higgs, ACTFL Foreign Language Education Series, No. 15 (Lincolnwood, IL: National Textbook Co., 1984), p. 44.

[16] The papers were published in the September, 1984 issue of *Foreign Language Annals.*

[17] Barbara F. Freed, "Establishing Proficiency-Based Language Requirements," *ADFL Bulletin,* 13, No. 2 (1982), 6-12.

[18] Theordore V. Higgs and Ray T. Clifford, "The Push Toward Communication," in *Curriculum, Competence, and the Foreign Language Teacher,* ed. Theodore V. Higgs, ACTFL Foreign Language Education Series, No. 13 (Lincolnwood, IL: National Textbook Co., 1982), p. 60.

[19] John L.D. Clark, *Foreign Language Testing: Theory and Practice* (Philadelphia: The Center for Curriculum Development, 1972), p. 5.

[20] See Marianne L. Adams, "Measuring Foreign Language Proficiency: A Study of Agreement Among Raters," in *Direct Testing of Speaking Proficiency: Theory and Application,* ed. John L.D. Clark (Princeton, NJ: Educational Testing Service, 1978), pp. 129-50.

[21] Michael Canale, "Considerations in the Testing of Reading and Listening Proficiency," *Foreign Language Annals,* 17 (1984), 349-57.

[22] See Omaggio, pp. 52-54, for some illustrative examples. She presents two sample exercises drawn from college French texts. Both exercises drill the formation of the present subjunctive, one in a series of unconnected sentences and the other in the context of a thematically unified group of sentences.

[23] The following list has been adapted from two sources: Bert F. Green, "Adaptive Testing by Computer," in *Measurement, Technology, and Individuality in Education,* ed. Ruth B. Ekstrom, New Directions for Testing and Measurement, No. 17 (San Francisco: Jossey-Bass, 1983), pp. 10-11; and Howard Wainer, "On Item Response Theory and Computerized Adaptive Tests,"

The Journal of College Admissions, 28, No. 4 (1983), 15.

[24] Published by D.C. Heath; Boston, MA.

[25] Floy Miller and Charlotte Cole, "Curriculum Project: NEH Institute, Haverford College, Summer, 1983" (unpublished mimeo).

[26] Omaggio, pp. 54-55.

[27] W. James Popham, "Measurement as an Instructional Catalyst," in *Measurement, Technology, and Individuality in Education,* ed. Ruth B. Ekstrom, New Directions for Testing and Measurement, No. 17 (San Francisco: Jossey-Bass, 1982), p. 22.