

A New Direction in Language Testing: Concern for the One

Harold S. Madsen

Doth he not leave the ninety and nine, . . . and seeketh that which is
gone astray? (Matthew 18:12)

An impressive LDS filmstrip for teachers entitled "The One" reminds us of the pitfall in thinking too exclusively of the quorum or class and in a story line about a handicapped girl focuses on the uniqueness of each individual. The implication is that reaching the one requires, for many of us, a new perspective. In the area of second-language instruction and testing, a combination of influences is beginning to provide such a perspective in an increasing concern for the individual student. This is reflected both in the professional literature and in our Church institutions as well.

INFLUENCES GENERATING CONCERN FOR THE INDIVIDUAL

Psychological Studies

Cognitive-style research attempts to identify the various ways individuals conceptualize and structure their environment;¹ and closely related to this are investigations of learning-style differences including modality preference (such as aural and visual), tempo (ranging from reflective to impulsive), and problem-solving strategies.² Varying learning styles have been found to differ in efficiency³ and to relate in some ways to differences in personality.⁴ Recent studies reveal a need for "alternate methods to match the educational

Harold S. Madsen, Brigham Young University professor of Teaching English as a Second Language (TESL), has published widely in the field of language testing and affect.

¹Kenneth M. Goldstein and Sheldon Blackman, *Cognitive Style: Five Approaches and Relevant Research* (New York: John Wiley and Sons, 1978).

²Len Sperry, *Learning Performance and Individual Differences: Essays and Readings* (Glenview, Ill.: Scott Foresman and Company, 1972), pp. 317-18.

³Lyle E. Bourne et al., *The Psychology of Thinking*, ed. James J. Jenkins (Englewood Cliffs, N. J.: Prentice-Hall, 1971), pp. 321-23.

⁴Henry C. Ellis, *Fundamentals of Human Learning, Memory, and Cognition*, 2d ed. (Dubuque, Iowa: William C. Brown Company, Publishers, 1978), pp. 284-85.

cognitive styles of different students''—as many as five or six for a heterogeneous group of about thirty students.⁵ Such studies appear also to have implication for test writers. One small example is the discovery that persons differing in learning style likewise vary in their ability to handle narrowly-spaced and broadly-spaced multiple-choice options on exams.⁶ In brief, current research supports the idea that individuals are unique in both their perceptions and learning styles.

The humanistic movement in psychology and language teaching also has contributed to this new emphasis on the individual. One reason is that humanists are interested in the total person, not simply the intellect. As in the voluminous literature on motivation,⁷ the interaction of emotions and intellect is closely examined. It has been shown, for example, that learning is significantly enhanced when students see the relevance of what they are studying to their personal lives.⁸ In fact, Rapaport holds that memory is intimately related to the emotional response of the learner.⁹ Humanistic instruction therefore strives to provide a blend (or "confluence," as educators label it) of the cognitive and the affective. One manifestation of this is communicative competence instruction, so fashionable this decade, in which the emphasis has shifted from mere linguistic accuracy to verbal exchanges that are socially appropriate, relevant, usually true, and hopefully of some importance to those communicating. Since humanistic education is concerned not only with increased language

⁵Renate A. Schulz, "Many Learners—Many Styles: The 1977 Central States Conference," *The Modern Language Journal* 41 (September–October 1977): 260. See also Lorayne B. Clavijo, "Implications of Individual Differences in ESL: An Evaluation and Pilot Study of Attribute–Treatment Interactions in the Pronunciation Class" (Major paper for TESL Certificate, BYU, 1979), p. 60.

⁶Sperry, *Learning Performance*, p. 111. Complementing the research being conducted on cognitive styles, learning styles, and personality styles are related investigations into instructional styles (see Frederic J. Bosco and Robert J. DiPietro, "Instructional Strategies: Their Psychological and Linguistic Bases," in *Toward a Cognitive Approach to Second-language Acquisition*, ed. Robert C. Lugton and Charles H. Heinle [Philadelphia: The Center for Curriculum Development, 1971], pp. 31–32). In addition, the interaction between learning styles and teaching styles is being systematically studied. A positive interaction is predicted when the learners encounter an approach matching their own characteristics (see Ellis, *Fundamentals of Human Learning*, p. 288).

⁷Robert C. Gardner and Wallace E. Lambert, "Motivational Variables in Second Language Acquisition," *Canadian Journal of Psychology* 13 (December 1959): 266–72; Wallace E. Lambert et al., "Attitudinal and Cognitive Aspects of Intensive Study of a Second Language," *Journal of Abnormal and Social Psychology* 66 (1963): 358–69; Robert C. Gardner, "A Survey of Attitudes of Educators of American Indian Children," *Research Bulletin* 66 (University of Western Ontario, 1968), referenced in John W. Oller, Jr., and Jack C. Richards, eds., *Focus on the Learner: Pragmatic Perspectives for the Language Teacher* (Rowley, Mass.: Newbury House Publishers, 1973), pp. 241, 245; Bernard Spolsky, "Attitudinal Aspects of Second Language Learning," *Language Learning* 19 (December 1969): 272–83; Earl W. Stevick, *Memory, Meaning, and Method* (Rowley, Mass.: Newbury House Publishers, 1976); Ellen Bialystok and Maria Frohlich, "Aspects of Second Language Learning in Classroom Settings," *Working Papers on Bilingualism* 13 (Toronto: Ontario Institute for Studies in Education, 1977).

⁸Beverly Galyean, "Language from Within: A Handbook of Teaching Strategies for Personal Growth and Self Reflection in the Language Classes" (Santa Barbara, Calif.: Confluent Education Development and Research Center, 1976); see also Stevick, *Memory, Meaning and Method*, pp. 38–40.

⁹D. A. Rapaport, *Emotions and Memory*, 5th ed. (New York: International Universities Press, 1971), p. 270.

proficiency but also with the many facets of personal growth,¹⁰ its content ranges from the academic to exploration of values, development of aesthetic sensitivity, improvement of self-image, and achievement of a greater capacity to relate effectively with others.

During this decade there has been a dramatic increase of interest in value exploration—much of it well before Watergate.¹¹ Along with this, there has been a growing awareness of the limitations of measurable behavioral objectives (still prominent in the Competency-Based Teacher Education Movement) as well as the limitations of experimental research and of language examinations themselves.¹² This perspective recalls the factors that transcend acquisition of knowledge or skills, in a variety of disciplines, ranging from the psychological readiness required in an athletic contest to the inspiration sought for in composing a piece of music.¹³

¹⁰Beverly Galyean, "Humanistic Education: A Mosaic Just Begun. An Integrative Approach to Foreign Language Teaching: Choosing among the Options," *ACTFL Foreign Language Education Series* 8 (1976).

¹¹Press accounts point up public concern that schools give attention to morals, values, and ethics. A 1974 survey showed that the number of school curriculum packages on moral education had more than doubled within the previous two years and that the number of graduate school dissertations on moral education had jumped from a total of 80 written during the three decades prior to 1970 to an average of 150 per year between 1970 and 1974. (*Christian Science Monitor* 1974. See also Renee Disick and Laura Barbanel, "Affective Education and Foreign Language Learning," *ACTFL Review, The Challenge of Communication* [1974]; Leland W. Howe and Mary Martha Howe, *Personalizing Education: Values Clarification and Beyond* [New York: Hart Publishing Co., 1975]; Sidney B. Simon et al., *Values Clarification: A Handbook of Practical Strategies for Teachers and Students* [New York: Hart Publishing Co., 1972]; Eleanor S. Morrison and Mila Underhill Price, *Values in Sexuality: A New Approach to Sex Education* [New York: Hart Publishing Co., 1974]; Mary S. Taylor, "Values Clarification in EFL," unpublished document, available through ERIC.)

¹²Typical of the many who oppose the imposition of performance objectives, Hans Guth reasons that this "asks them for 'data' at a time when many of them are in search of 'soul.' It asks them to make their students 'perform' when many of them are concerned with 'reaching' the student. It asks them to administer tailor-made 'learning sequences' at a time when many are concerned with liberating the students' locked-in creative and human potential." (Hans P. Guth, "The Monkey on the Bicycle: Behavioral Objectives and the Teaching of English," *The English Journal* 59 [1970]: 785–86.) In the same vein, Rothstein, arguing for a humanistic approach to goal setting, points out that a quiet student listening to a discussion of Frost's "Birches" might be affected at some later time; another student could present a clear analysis but might not ever be influenced by the poem. He concludes that behavioral objectives, far from constituting a progressive step in education, are "one of the most reactionary developments to be employed in recent times." (Herbert M. Rothstein, "A Humanistic Approach to Behavioral Objectives," *The English Journal* 60 [1971]: 760–61; for a survey of pros and cons, see Harold S. Madsen, "Achieving Certification through a Modified Competency-Based TESL Teacher Education Program," *TESOL Quarterly* 9 [December 1975]: 353–65.) Reflecting this collective concern, the National Council of Teachers of English expressed itself as follows: "Resolved, That those who purpose to employ behavioral objectives be urged to engage in a careful appraisal of the possible benefits and the present limitations of English with reference to the humanistic aims which have traditionally been valued in this discipline. And be it further Resolved, That those in the profession who do undertake to write behavioral objectives (a) make specific plans to account for the total English curriculum; (b) make an intention to preserve (and, if need be, fight for) the retention of important humanistic goals of education; and (c) insist on these goals regardless of whether or not there exist instruments at the present time for measuring the desired changes in pupil behavior [emphasis added]. (*The English Journal* 59 [1970]: 501.) Also, while experimental research has provided important insights regarding major differences in individual cognitive style, a substantial portion of the theory undergirding humanistic education stems from "data provided by subjective experiences" (Galyean, "Humanistic Education"; for an exploration of examination concerns, consider Paul L. Houts, ed., *The Myth of Measurability* [New York: Hart Publishing Co., 1977]; also John W. Oller, Jr., and Kyle Perkins, eds., *Language in Education: Testing the Tests* [Rowley, Mass.: Newbury House Publishers, 1978].)

¹³Brahms, for example, considered inspiration such a vital component in the creative process that he felt an atheist could never produce a truly great masterpiece (Reid N. Nibley, forum address, BYU, 9 October 1979).

Underpinning humanistic educational concepts is Maslow's humanistic psychology with its principle of self-actualization.¹⁴ But even more prominent in language teaching throughout this decade are concepts from cognitive psychology and transformational grammar. Rejecting the behaviorists' position that learning is simply a matter of conditioning or the "formation of habits through responses to outside stimuli," proponents of cognitive psychology espouse a mentalist theory, indicating that effective language acquisition comes from meaningful learning and that it is a rule-governed, creative process.¹⁵

Broad acceptance of such concepts has contributed to the shift in emphasis from the teacher to the learner.

Shifts in Language Teaching Methodology

Another influence has been the trend away from traditional audio-lingual methodology with its heavy emphasis on repetitive drills designed to develop automatic responses and language habits. In its place have appeared cognitive modifications as well as cognitively-oriented methodologies that attempt to tap the learner's innate capacity to acquire language. Lozanov's Suggestology attempts to increase fluency and the rate of learning through "hypernesia" by removing anxiety-producing psychological barriers.¹⁶ Gattegno's Silent Way seeks to promote self-reliance and personal initiative by reducing the teacher's verbal input and allowing time for reflection.¹⁷

Perhaps the best example is Curran's Counseling Learning which, like humanistic education, aims at strengthening the entire individual by developing learner initiative, interaction among students, and mutual respect between teachers and learners as well as increased self-worth.¹⁸ And a teaching approach developed by

¹⁴Abraham H. Maslow, *Motivation and Personality* (New York: Harper & Row, 1954); Abraham H. Maslow, *Toward a Psychology of Being*, 2d ed. (New York: Van Nostrand, 1962); Abraham H. Maslow, *Eupsychian Management* (Homewood, Ill.: Irwin-Dorsey, 1965); Frank G. Goble, *The Third Force: The Psychology of Abraham Maslow* (New York: Grossman Publishers, 1970).

¹⁵For a classical statement of the contrasts between mentalist and behaviorist theory, see Dwight Bolinger, "The Theorist and the Language Teacher," *Foreign Language Annals* 2 (October 1968): 30-41; see also Kenneth Chastain, *Developing Second Language Skills: Theory to Practice*, 2d ed. (Chicago: Rand McNally College Publishing Co., 1976).

¹⁶Georgi Lozanov, *Suggestology* (Sofia, Bulgaria: Nauka Izkistvo, 1973); see also Harold S. Madsen, "Innovative Methodologies Applicable to TESL," in *Teaching English as a Second or Foreign Language*, ed. Marianne Celce-Murcia and Lois McIntosh (Rowley, Mass.: Newbury House Publishers, 1979).

¹⁷Caleb Gattegno, *The Common Sense of Teaching Foreign Languages* (New York: Educational Solutions, 1976).

¹⁸Charles A. Curran, *Counseling Learning: A Whole-Person Model for Education* (New York: Grune and Stratton, 1972); Earl W. Stevick, "Counseling Learning: A Whole-Person Model for Education," *Language Learning* 23 (December 1973): 259-71; and Madsen, "Innovative Methodologies."

Walter Gong of San Jose State University and promoted at Brigham Young University, particularly in the College of Humanities, is designed to educate an individual in the broadest sense:

We must lead the student into the application of what he has been taught to the realities of life and the expanding of the central point by the relation of it to everything that he already knows and has experienced. Our job as teachers cannot end with the mere teaching of facts; our job cannot end until we have led our learners to relate our facts to principles and to the total organisms that are themselves. Thus their lives will be blessed by growth and movement toward understanding; and the lives of their wives, husbands, children, and friends will be blessed by an expansion of comprehension.¹⁹

With emphases such as these, contemporary methods and approaches are unmistakably contributing to our contemporary focus on the learner.²⁰

NATIONAL TRENDS IN LANGUAGE TESTING

Preparing Tests to Meet Special Needs

How, then, does this translate into language evaluation? Basically, we have moved in the direction of lifelike language in a test form designed for nonnative speakers of the language.

A broad brush stroke representation of language testing trends in recent decades shows it moving in the 1940s and 1950s from a long period of largely intuitive test making (ranging from grammatical parsing and labeling to translation, essay, and précis writing) into a scientific era, during which "less attention was paid to *what* was tested than to *how* it was tested."²¹ While earlier tests often required analysis of the language, the newer objective measures typically required mastery of discrete segments of the language, such as the correct grammatical phrase, vowel sound, or lexical item in a series of unrelated sentences. But now during the communicative era of testing, we generally seek fuller contextualization and a closer approximation of real-life situations. This means an increased use of tests

¹⁹Edward L. Hart, "The Gong Approach," in unpublished handout, "Some Common Sense on Learning How to Teach," by Eugene England, 1979, p. 4.

²⁰It is no accident that current titles in language periodicals and texts so frequently reflect this concept: e.g., Charles A. Findley, "Focus on the Learner: Security, Alienation and Risk-taking," *RELC Journal* (Singapore) 9 (1978): 69-76; Diane W. Birckbichler and Alice C. Omaggio, "Diagnosing and Responding to Individual Learner Needs," *The Modern Language Journal* 62 (November 1978): 336-45; Oller and Richards, eds., *Focus on the Learner*.

²¹Randall L. Jones, "Testing: A Vital Connection," in *The Language Connection: From the Classroom to the World*, ed. June K. Phillips (Skokie, Ill: National Textbook Company, 1977), p. 238.

measuring actual performance such as conversational competence or note taking; these involve simultaneously not only the processing of grammar, lexis, and phonology but also coping with fluency, semantic sensitivity and possibly social appropriateness as well. While multiple-choice language exams still exist, test writers tailor these carefully for nonnative speakers. For example, distractors are drawn from foreign student compositions or exercises, and native-speaker errors (such as, "she might of told him") are avoided.

Reducing redundancy in a prose passage has been found to be a very effective device in differentiating between native and nonnative language ability. Experimental forms include a noise test consisting of dictated sentences overlaid with "white" or "pink" electronic noise that partially obscures inflections.²² Another is "gapped listening"—a tape-recorded reading or news broadcast from which portions are deleted at regular intervals. Students take notes and then answer questions based on the original tape. A third is the *Integrative Grammar Test*—dictated sentences replete with assimilation, contraction, and reduction. Students are asked to write out the full form of the second word they hear in each item. For example, after listening to "Wouldja like-im to help-ya?" they would write *you* in the blank.

In addition to those experimental forms previously mentioned, there are other reduced redundancy tests in wide use: One is the traditional dictation, now used with one or two modifications. The passage is first read without pauses and at normal speed. Then it is read a second time at normal speed but with pauses for students to write down what they have heard. Very important is the length of the phrases dictated—about five to nine words per phrase group. A brisk third reading without pauses provides for some proofreading. Punctuation can be given and misspellings ignored. To provide consistency in scoring, all errors are weighted the same. More popular still is the cloze test. This powerful instrument consists of a prose passage from which words are deleted at random intervals, typically every seventh word. From the remaining context, students are required to

²²Both "white" and "pink" noise can be described electronically and metaphorically: "White noise sounds like this: sh/sh/sh/sh/sh. It's simply random frequencies at random amplitudes, the basic kind of noise that you hear in back of radio broadcasts. It's called white because it has the same characteristics as white light, that is, all frequencies are represented at random. I guess pink noise is just a little more regular in frequency." (Randall L. Jones and Bernard Spolsky, *Testing Language Proficiency* [Arlington, Va.: Center for Applied Linguistics, 1975], p. 70.) White noise is generated electronically utilizing the same amount of energy per cycle segment; pink noise, also generated electronically, utilizes the same amount of energy per octave. White noise is a hissing sound like escaping air from a tire; pink noise is a low rumbling sound like the wind noise in a microphone (and in the same frequency as the speech range).

supply the missing words, as one does automatically in real-life situations, for example, when conversing in a noisy department store or in an air terminal. Variations include selected deletion (with only function words omitted, for example) or multiple-choice cloze. The dictation as outlined here and particularly the cloze procedure provide excellent measures of general language proficiency.²³

In addition to tailoring language exams for nonnatives, test writers are now fashioning a rich variety of evaluation instruments to accommodate the many limitations and special objectives of second-language learners. For children, who would probably be intimidated by or unable to cope with standard paper-and-pencil tests, there are attractive picture tests that can be individually administered in a relaxed conversational manner.²⁴ Many of these can also be used as bilingual tests to assess language dominance. For adults with limited language skills, one can use an oral test with picture cues (for example, the *Ilyin Oral Interview*²⁵) or such measures as a listening test with printed *native-language* options.²⁶ Besides standard reading tests, multiple-choice cloze is now available for evaluating a skill like reading, when only passive recognition is required.²⁷ For advanced students seeking admission to American universities, there are sophisticated test batteries such as the TOEFL or MTELP (*Test of English as a Foreign Language* and *Michigan Test of English Language Proficiency*). For highly trained civil servants whose oral communication skills need evaluating, there is the remarkably adaptable FSI test (*Foreign Service Institute Oral Interview*). And there are tests for language acquisition research,²⁸ for translators (*Translation Evaluation Program*), and even for *group* evaluation of student speaking skills.²⁹

²³John W. Oller, Jr., "Dictation as a Device for Testing Foreign Language Skills," *English Language Teaching* 25 (June 1971): 254-59; John W. Oller, Jr., and Christine Conrad, "The Cloze Procedure and ESL Proficiency," *Language Learning* 21 (December 1971): 183-96; and scores of other articles written throughout this decade.

²⁴Marina K. Burt, Heidi C. Dulay, and Eduardo Hernandez Ch., *Bilingual Syntax Measure* (New York: Harcourt Brace Jovanovich, 1975); Edward A. DeAvila and Sharon E. Duncan, *Language Assessment Scales I [LAS]* (Corte Madera, Calif.: Linguametrics Group, 1976).

²⁵Donna Ilyin, *Ilyin Oral Interview* (Rowley, Mass.: Newbury House Publishers, 1972).

²⁶Rebecca M. Valette, *Modern Language Testing*, 2d ed. (New York: Harcourt Brace Jovanovich, 1977); Harold S. Madsen, "An Indirect Measure of Listening Comprehension," *The Modern Language Journal* 63 (December 1979): 429-35.

²⁷D. Porter, "Modified Cloze Procedure: A More Valid Reading Comprehension Test," *English Language Teaching Journal* 30 (January 1976): 151-55.

²⁸Ann Fathman, "The Relationship between Age and Second Language Productive Ability," *Language Learning* 25 (December 1975): 245-53; see also Stephen D. Krashen, S. V. Sferlazza, and Ann Fathman, "Adult Performance on the SLOPE Test: More Evidence for a Natural Sequence in Adult Language Acquisition," *Language Learning* 26 (June 1976): 145-51.

²⁹David Folland and David Robertson, "Towards Objectivity in Group Oral Testing," *English Language Teaching Journal* 30 (January 1976): 156-67.

The Concern over Bias

Perhaps the most dramatic concern for the individual is manifested in the widespread desire to eliminate all forms of bias from educational tests, particularly language exams. Reflecting older complaints that IQ tests measure language maturity more than native intelligence, various critics are assailing the widespread bias in standardized educational tests. Houts, for example, marshals a formidable array of educators and researchers who assert that present-day educational tests exacerbate social inequities by stigmatizing children instead of reducing these inequities by encouraging a modification of the curriculum to "meet variations in interests, talents, backgrounds."³⁰ In their meticulously documented book, Oller and Perkins together with other researchers note the content similarity in achievement batteries, intelligence tests, personality inventories, and language proficiency tests. Then, through statistical analysis, they demonstrate that language proficiency is "a major variable" in the tests evaluated (accounting for .58 to .88 of the variance in the subscores of the *California Achievement Test*, for example). These writers conclude that, even for native speakers, tests that purportedly measure intelligence or achievement or personality may primarily be measuring language proficiency. Nonnative speakers are therefore in real danger of being improperly evaluated by such instruments:

It is obvious that the student who speaks and understands the language variety of the test will have an advantage over the student who is more familiar with a different variety. That is, the tests are clearly biased against speakers of non-majority varieties of English.³¹

So great has the concern become over possible inequities of this kind that legal actions have been taken to protect the rights of minority groups in America. The most dramatic was the 1974 U.S. Supreme Court decision in the *Lau v. Nichols* case. The Court decided in favor of a class action suit filed on behalf of 1,800 Chinese-speaking students in San Francisco who were allegedly denied equal educational opportunities. Rejecting the school board's "English only" policy, the majority opinion reasoned that under California's state-imposed standards—

³⁰Houts, *Myth of Measurability*; John N. Leach, "Bias in Standardized Testing: An Update" (Paper presented at the TESOL Convention, Boston, March 1979); and Eugene J. Briere, "Cross-Cultural Biases in Language Testing," in *Focus on the Learner*, ed. Oller and Richards.

³¹Oller and Perkins, *Language in Education*, p. 34; see also pp. 33, 94.

There is no equality of treatment by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education. . . . Imposition of a requirement that, before a child can effectively participate in the educational program, he must already have acquired those basic [English] skills is to make a mockery of public education.³²

Congress subsequently funded nine General Assistance Centers to help schools meet the needs of students with limited or no proficiency in English. As a result, public schools are now being assisted to provide unbiased evaluation of nonnative English speakers. And appropriate testing has become an important concern, particularly language dominance assessment and evaluation design.

Similar actions preceded and followed the Lau case. In 1972, the Supreme Court focused on testing that places students in classes for the educable mentally retarded (EMR). The Court determined that the tests were culturally and socio-economically biased, and it ordered that the students be evaluated by unbiased instruments. When retested with unbiased instruments, two-thirds of the Black students tested out of EMR into regular classes.³³ Also recognizing the limited language proficiency of many immigrants and minority groups, Title VII of the Civil Rights Act of 1964 "mandated employers to cease the use of test results to discriminate among potential employees if and when job relatedness could not be established."³⁴ In recent years, court cases and legislation on tests with language bias have increased. These actions are intended to eliminate bias in a variety of ways, such as freeing children of faulty and stigmatizing labels (for example, "mentally retarded") or halting the practice of educational "streaming" into inappropriate programs. But the broader purpose of these legal actions related to language testing is to preserve individual rights not only in the area of citizenship and voting but also in the realms of education, employment, and human dignity.³⁵

Cultural bias of another sort is being examined by educators. A Korean teacher of English recently disclosed that in 1976 he took the prestigious TOEFL (*Test of English as a Foreign Language*), for those seeking to enter an American university. On the listening section, he

³²Lau v. Nichols, 414 U.S. 566 (1974).

³³Patricia J. Nakano, "Educational Implications of the *Lau v. Nichols* Decision," *Viewpoints on English as a Second Language*, ed. Marina Burt et al. (New York: Regents Publishing Company, 1977), pp. 221-22.

³⁴Gilbert N. Garcia, "Testing in the Second Language: Mandates and Decrees" (Paper presented at the TESOL Convention, Boston, March 1979).

³⁵Gilbert N. Garcia, "Three Wrongs Can Make a Wrong" (TESOL Summer Institute Lecture, UCLA, Los Angeles, 24 July 1979).

was asked questions related to a passage on American literature. Having studied this subject extensively, he performed extremely well, receiving a total score of 623 (a score of 500 would have admitted him into most American universities). After teaching English for two years, he once again took the TOEFL. But this time he encountered a passage on European literature and another on a scientific topic. Though well grounded in Oriental history and the arts, he knew little about the two exam subjects. Consequently, he scored significantly lower than the first time, receiving a total of only 580, or a drop of close to 40 points.³⁶ Since the TOEFL is a language proficiency test and not a test of general knowledge, such fluctuations reveal a potential cultural bias of alarming proportions.

Also it has been disclosed recently that certain reading comprehension tests evaluate cultural knowledge as well as language proficiency and "thereby discriminate against ESL (English as a Second Language) students"; several tests contain ten to fifteen percent biased items.³⁷ Consider the following examples from various reading tests:

"There are red and white stripes and white stars in our flag. Our flag contains one . . . for every state.

(a) stripe (b) star . . .

"The French regarded potatoes like most Canadians regard:

(a) spinach (b) tomatoes (c) horsemeat (d) margarine

[note: In the story the French dislike potatoes.]

"Sam won at marbles because he could . . . straighter than Bill.

(a) show (b) shoot (c) draw (d) run

"The Yankee peddler traded as far west as the Mississippi and as far south as Louisiana. He operated . . .

(a) over most of the country (b) as far south as Louisiana . . .

"Pam went to the party with a tall pointed black hat, long black cape and a broom. She was dressed as a . . .

(a) witch (b) ghost (c) cowgirl (d) pumpkin

"Bill ran out on his front porch to watch the firetruck. He lives in

(a) a big apartment (b) a city house (c) a trailer . . ."³⁸

³⁶Oryang Kwon, personal interview, Provo, Utah, February 1980.

³⁷Bernard Mohan, "Cultural Bias in Reading Comprehension Tests," in *On TESOL '79: The Learner in Focus*, ed. Carlos A. Yorio (Washington, D.C.: Teachers of English to Speakers of Other Languages, 1979), pp. 171-77.

³⁸*Ibid.*

An illustration is provided by a prominent language-testing specialist who evaluated the *Wepman Auditory Discrimination Test*:

As a part Mohawk Indian, I have been trained never to trust a stranger behind my back in a small room. This sort of thing is frequently true of American Indians, Blacks, and Chicanos. Since the test administrator (tall, big, authoritative, and threatening) stands *behind* the children during the test (presumably to avoid lip cues), I wonder how valid the results are with groups who have been culturally trained *not* to allow a threatening figure to get behind them?³⁹

Very recently test bias has even been found in the very form that language tests take, specifically in integrative and discrete-point formats. We recall that integrative tests require the processing of several language components simultaneously, as on dictation or essay tests, while a discrete-point test focuses on one language component at a time, as on a multiple-choice grammar test. Farhady suggests that all-discrete-point or all-integrative tests discriminate against students of some nationalities. For example, in one study, students from Israel and France scored significantly higher on an integrative placement test than on a parallel discrete-point test. But students from Taiwan and Korea scored significantly higher on the discrete-point form. Farhady also suggests the possibility of a sex bias related to test form.⁴⁰

Still another concern is that of eliminating random bias related to the administration and scoring of language examinations. For instance, on oral tests, examiners that speak the native language of an examinee sometimes subconsciously overlook certain errors simply because they encounter these so frequently in the classroom, while errors made by others may be looked at more critically. Moreover, listening tests administered in large rooms where the sound reverberates can result in a weaker showing than normal for students with even minor hearing disabilities. And a number of researchers have investigated the techniques for scoring cloze tests that provide the most valid measure of general language proficiency.⁴¹

³⁹Briere, "Cross-Cultural Biases," p. 219; see also William Labov, "Academic Ignorance and Black Intelligence," *Atlantic Monthly* (June 1972): 59-67.

⁴⁰Hossein Farhady, "Discrete-point vs. Integrative Tests," unpublished paper, 1979; and also Hossein Farhady, "Test Bias in Language Placement Examination," in *On TESOL '79*, ed. Yorio et al., pp. 162-70.

⁴¹Charles J. Alderson, "The Cloze Procedure and Proficiency in English as a Foreign Language," *TESOL Quarterly* 13 (June 1979): 219-27; James D. Brown, "Correlational Study of Four Methods for Scoring Cloze Tests" (Paper presented at the TESOL Convention, Boston, March 1979); and Karen H. Mullen, "An Alternative to the Cloze Test," in *On TESOL '79*, ed. Yorio et al., pp. 187-92.

LANGUAGE TESTING IN THE CHURCH

In a variety of interesting ways, language testing in the institutions of The Church of Jesus Christ of Latter-day Saints reflects or transcends secular concerns for the individual. Testing is seen not only as a means of evaluation but also as an aid to motivation and personal improvement. For example, in one Church operation, posters have blossomed recently which quote Elder Thomas S. Monson's observation that "when performance is measured, performance improves. When performance is measured and reported, the rate of improvement accelerates."⁴²

Language Interests for Special Needs in the Church.

In addition, language tests are being developed for more specialized purposes than ever before. BYU utilizes, of course, the usual EFL tests that screen foreign-language applicants to the university, plus tests for placing matriculated students needing further English training and tests for nonmatriculated people in intensive ESL courses.⁴³ Besides these, the university has prepared examinations in twenty-five foreign languages to enable returned missionaries and others to turn language skills into college credit.⁴⁴ Specialized tests are administered not only to missionaries returning from non-English speaking countries but also to prospective missionaries and to missionaries in training. A commercially prepared language aptitude battery⁴⁵ is taken by all persons being processed for full-time missions. The Brethren have access to these scores for reference when making missionary assignments.

Missionaries in training are evaluated on five specialized measures. Their foreign language mastery is assessed by means of the

⁴²Thomas S. Monson, translation evaluation program poster in Language and Intercultural Resource Center (LIRC), BYU.

⁴³Tests that screen university applicants include the TOEFL and the MTELP as well as the ALIGU, CELT, GCE, etc. Those placing foreign students at the appropriate level include the BYUEPT and the IGT; the *Michigan Placement*, *New Horizons* and *Win Locator* are used for grouping nonmatriculated students. And the BYUEB is used to screen graduate foreign language applicants to the BYU ESL program.

⁴⁴Students can acquire up to sixteen hours of foreign-language credit on a pass/fail basis with graded credit only at the 201 level. Enrollment in an appropriate upper division course permits the usual letter grade to be assigned for all hours. Students can count this credit toward their General Education Category III requirements. In addition to tests in the usual European and Asian languages, the *Foreign Language Achievement Test Series* is also administered in languages such as Afrikaans, Aymara, Cakchiquel, Farsi, Indonesian, Samoan, Serbo-Croatian, Tahitian, Tongan, Hebrew, and Thai. These tests are administered to BYU and also non-BYU students if approved by the other institution. (Deborah L. Coon, project coordinator, LIRC, BYU, personal interview, approximately 25 September 1979.)

⁴⁵The test used is the *Modern Language Aptitude Test* (MLAT). It is administered weekly at the Missionary Training Center and by stake presidents in outlying areas. The MLAT battery is designed to predict success in learning a foreign language.

U.S. government's prestigious FSI oral examination.⁴⁶ The *Discussion Mastery Test* (DMT) is administered periodically to check the missionaries' progress with the basic missionary discussions.⁴⁷ This is complemented by the *Teaching Skills Evaluation* (TSE) that looks at the effectiveness with which the gospel discussions are presented, focusing on the mechanics of the presentation, teaching effectiveness, and missionary spirit. A fourth evaluation is the *Final Scriptures Score*, tallying the number of discussion-related scriptures that have been mastered. And finally, there is the *Speak-Your-Language* score, a cumulative self-rating that is recorded daily on a four-point scale; this indicates the degree to which the missionary-in-training uses the target language in daily living.

This elaborate evaluation at the Missionary Training Center is intended to provide not only perspective and incentive for missionaries but also the means of evaluating training and materials.⁴⁸

The Church's Translation Division also has begun utilizing a variety of highly specialized language measures. One system of evaluation is used for prospective translators of emerging languages (languages in countries where the missionaries are just being introduced) and another for those seeking to translate in the established languages.⁴⁹ For translator applicants of emerging languages, a short test called the *Translator Screen* supplements a personal interview. This integrative test consists of three passages from Church literature that the prospective translator is asked to translate. Because

⁴⁶Missionaries at the Missionary Training Center (MTC), except those studying Serbo-Croatian, are administered the FSI test at the end of the fourth and eighth weeks. Possible scores range from 0 to 5, with 1 constituting "survival-level proficiency" and 5 essentially "native-speaker proficiency." Upon completing their MTC training, missionaries usually average about 1 to 1+. This points out that while the elders and sisters acquire a rich background in a highly specific religious register, they leave for the mission field with rather limited skills as far as general conversational ability is concerned. Nevertheless, their progress during their brief stint in the MTC is extremely good, by any standard. Ideally, MTC instructors are expected to have an FSI rating of at least 3 in the target language. While the MTC does not report FSI scores to missionaries, it does provide extensive feedback on the nature of each person's difficulties or strengths. At times, language skills have also been evaluated on a written grammar test. This has recently been suspended, however. (Lane D. Ward, assistant director of training, testing, and zone coordinators, MTC, personal interview, 2 October 1979; Eric Ott, instructional evaluator, MTC, personal interview, 25 October 1979; Cecilia Nihlen, FSI examiner, MTC, personal interview, 2 and 22 October 1979; and Allen C. Ostergar, Jr., director of training, MTC, personal interview, 1980.)

⁴⁷This twenty-five-minute test covering the eight discussions samples twenty of the fifty-six concepts that have been learned. Discussion pass-off is handled in pairs, with appropriate interaction expected between the two missionaries.

⁴⁸Eric Ott, instructional evaluator, MTC, personal interview, 1980.

⁴⁹Translators selected to translate into the varying number (now fifty-four) of emerging languages are presently evaluated less extensively than are those applying to translate into the eighteen established languages—partly because the former translators are often more scarce and partly because the scope of their translations is more limited. They translate packages of materials in phases, starting with Joseph Smith's testimony, the Gospel Principles manual, Book of Mormon selections, and some basic organizational guidebooks. Translators of established languages deal with material from the entire spectrum of Church operations. (Robert W. Bushman, coordinator of training, LDS Church Translation Division, personal interviews and correspondence, 4 and 8 October 1979, 25 June 1980. Also see Joseph G. Stringham, "The Church and Translation," *Brigham Young University Studies* 21 [1981]: 69–90.)

of limited personnel resources in emerging languages, the hiring supervisor, who is not skilled in the language of the prospective translator, checks the faithfulness of the translations by questioning the translator on specific points of the translation and by using back translations. In this manner, the supervisor gains an adequate impression of the translator's English comprehension skill, background knowledge, and translation skill. Once on the job, emerging language translators double check each other's work and can assess the appropriateness of a translation's language for its audience.

Translator applicants of the established languages are very carefully screened. They begin by taking a practice test to familiarize themselves with the format of the comprehensive translator test that follows. The latter consists of an English reading comprehension examination and a translation examination. The test of English comprehension includes a commercial vocabulary and reading exam (the *Iowa Silent Reading Test*) and a subtest on Church terminology, which simultaneously tests familiarity with Church history, doctrine, and policy. The translation examination requires actual translating of excerpts from Church literature.⁵⁰ In short, translators are expected to comprehend English well, have a good general background (ranging from Church doctrine and history to cultural understanding and world events), possess effective writing skills in their native language, and demonstrate good transfer skills by finding suitable equivalents in their native language for meanings expressed in English. In addition, they need to have a suitable temperament for their assignment. After a translator is hired, the quality of the translator's production translations is measured with the same instruments under a program of quality control. Measurements so derived are considered in such personnel management decisions as assignments, salary raises, and promotions.

Factors Complementing or Transcending Language Skills

As in the secular realm, language evaluation in the Church is regarded as an important undertaking—one that often requires a highly specialized design to meet specialized needs and interests. And as we have seen in academia nationally, Church institutions likewise recognize the limitations of such instruments. For example, at Church schools including BYU, admission is based not only on

⁵⁰Examinees translate six passages, 75–100 words each, from the following types of Church writings: general, technical, literary, poetic, and children's. Three judges independently rate the quality of each translated passage on the basis of faithfulness, clarity, mechanics, and overall value. Results are tabulated and analyzed by three reference groups: the local translation staff, other translators of the geographic region, and translators Churchwide.

academic preparation but also on moral worthiness.⁵¹ And in screening teachers for overseas Church schools, the Church Educational System is concerned not only with applicants' academic ability and moral character but also their adaptability and cultural sensitivity.

We can see parallels in the Church missionary program. While language aptitude scores are available on prospective missionaries, it is the inspiration of Church leaders that is the ultimate factor in determining the call. And we have seen that much of the evaluation made of those being prepared for foreign language missions pertains more directly to gospel principles than to language proficiency. We recall, for instance, that one of the criteria for evaluating missionaries' gospel teaching skills is their spirit. Even though language proficiency is important, the MTC recognizes that it is individual testimony that sparks conversion. While the language component can be measured with considerable accuracy, the most significant attribute of testimony is simply not quantifiable.

Testimony and understanding of gospel principles are also valued in Church translators. Moreover, the four personal qualities sought for in new translators are all extra-linguistic: ability to change, willingness to learn, ability to work with others, and capacity to accept criticism.⁵² Ultimately, then, selection of translators involves assessing not only specialized language skills but also very personal attributes related to gospel understanding and temperament.

It is not surprising, in the light of this discussion, that BYU offers formal coursework in language testing—courses designed to provide skill in developing the sophisticated examinations required nowadays as well as the ability to interpret test results and identify the limitations of language assessment. Moreover, interest in the individual examinee is evidenced in current or recently completed experimental research at BYU: testing studies have focused on the special language

⁵¹A confidential statement of the applicant's moral worthiness is provided by his bishop. A non-LDS applicant is also to be interviewed by a bishop, if one is available, or by a clergyman of his faith. In addition, the applicant commits himself in writing and in the presence of his church leaders as follows: "I hereby commit myself to do the following while enrolled at BYU if I am admitted or readmitted: (a) Conduct my personal life consistent with the standards of Christian living taught by The Church of Jesus Christ of Latter-day Saints *both on and off campus*; (b) Adhere to the other requirements of the attached Code of Honor and Dress and Grooming Standards as defined by the Board of Trustees" (BYU Application for Admission and Scholarships, Office of Admissions, BYU). Ethical and moral values receive attention alongside academic subjects. For example, prospective elementary school teachers take a course titled "Value Clarification in Education," and prospective secondary school teachers are evaluated in part on their ethical and professional behavior (BYU "Evaluation of Student Teaching Performance" form; Eldon H. Puckett, BYU Department of Elementary Education, personal interview, 11 October 1979; and Jeffrey M. Tanner, BYU director of admission, personal interview, 1 October 1979).

⁵²Robert W. Bushman, coordinator of training, LDS Church Translation Division, personal interviews and correspondence, 4 and 8 October 1979, 25 June 1980.

problems faced by American Indians,⁵³ Japanese,⁵⁴ Orientals,⁵⁵ and Swedish immigrants to the United States.⁵⁶ Other studies have attempted to meet the testing needs of those in developing countries where listening comprehension may be difficult to measure,⁵⁷ and of those with very limited second-language proficiency.⁵⁸ Finally, some of the most interesting test research related to individual needs is that which studies "affect"—each examinee's emotional reaction to various exam formats.⁵⁹

Measurements taken include student perceptions of exam difficulty, level of performance, fairness, relevance to instruction, validity, and amount of anxiety or frustration experienced. These are analyzed in reference to exam type and examinee background. The results of this kind of research promise not only to help improve the accuracy of language measurement but also to reduce test frustration and thereby contribute to an improved climate for language acquisition.

In conclusion, recent secular concern for the total individual, including his or her social-emotional fulfillment, has resulted in language evaluation designed to meet highly specific needs as well as an increased awareness of the limitations inherent in educational assessments and language tests. Understandably, we find in Church institutions a similar concern for the individual. Along with efforts to devise specialized language examinations is an awareness that linguistic skills must be matched by personal and spiritual attributes and that respect for the intellect must be matched by respect for the worth of a soul.

⁵³Patricia B. Dyck, "Written Usage of American Indian and Anglo College Students: A Comparative Error Analysis" (M.A. thesis, BYU, 1979).

⁵⁴Chie Nishimura, "An Analysis of the Uses of Articles by Japanese Students" (Major paper for the TESL Certificate, BYU, 1980).

⁵⁵Setsuko Shimizu, "The Difficulties for Oriental Students in Learning English Articles" (Major paper for the TESL Certificate, BYU, 1980).

⁵⁶Cecilia Nihlen, "Attained English Proficiency by First Generation Swedish Immigrants in Salt Lake City" (Major paper for the TESL Certificate, BYU, 1980).

⁵⁷Herbert Leon Twyman, "Validation of the Alternate Modality Listening Exam (AMLEX): Studies with a Heterogeneous Group of Second Language Learners" (M.A. thesis, BYU, 1979).

⁵⁸Sheila Maluf, "The Use of Native Language Cues: Evaluating Foreign Language Listening Skills of Low Proficiency Students" (M.A. thesis, BYU, 1979); and Sheila Maluf and Harold S. Madsen, "Using Native Language Cues in Evaluating Low Proficiency Language Learners" (Paper presented at the Deseret Language and Linguistic Society Annual Symposium, BYU, 5 April 1979).

⁵⁹Diana G. Stevenson, "The Experimental Evaluation of Test Affect" (M.A. thesis, BYU, 1979), and Randall L. Jones, Harold S. Madsen, and Bruce L. Brown, "Evaluating Affective Variables in Second-Language Testing" (Paper presented at the TESOL Convention, San Francisco, March 1980).