

A
Guide to
LANGUAGE
TESTING ■
Development
■
Evaluation
■
Research

■ **GRANT HENNING**

Preface by Chomsky

It is about half a century since the study of language undertook a rather new course, while renewing some traditional concerns that had long been neglected. The central change was a shift of attention from behavior and the products of behavior (texts, corpora, etc.) to the internal mechanisms that enter into behavior. This was part of a general shift of perspective in psychology towards what became known as “cognitive science,” and was in fact a significant factor in contributing to this development.

With this departure from prevailing structuralist and behaviorist approaches, the object of inquiry becomes a property of individual persons, my granddaughters for example. We ask what special properties they have that underlie an obvious but nonetheless remarkable fact. Exposed to a world of “buzzing, booming confusion” (in William James’s classic phrase), each instantly identified some intricate subpart of it as linguistic, and reflexively, without awareness or instruction (which would be useless in any event), performed analytic operations that led to knowledge of some specific linguistic system, in one case, a variety of what is called informally “English,” in another a variety of “Spanish.” It could just as easily been one of the Chinese languages, or an aboriginal language of Australia, or some other human language. Exposed to the same environment, their pet cats (or chimpanzees, etc.) would not even take the first step of identifying the relevant category of phenomena, just as humans do not identify what a bee perceives as the waggle dance that communicates the distance and orientation of a source of honey.

All organisms have special subsystems that lead them to deal with their environment in specific ways. Some of these subsystems are called “mental” or “cognitive,” informal designations that need not be made precise, just as there is no need to determine exactly where chemistry ends and biology begins. The development of cognitive systems, like others, is influenced by the environment, but the general course is genetically determined. Changes of nutrition, for example, can have a dramatic effect on development, but will not change a human embryo to a bee or a mouse, and the same holds for cognitive development. The evidence is strong that among the human cognitive systems is a “faculty of language” (FL), to borrow a traditional term: some subsystem of (mostly) the brain. The evidence is also overwhelming that apart from severe pathology, FL is close to uniform for humans: it is a genuine species property. The “initial state” of FL is determined by the common human genetic endowment. Exposed to experience, FL passes through a series of states, normally reaching a relatively stable state at about puberty, after which changes are peripheral: growth of vocabulary, primarily.

As far as we know, every aspect of language — sound, structure, meanings of words and more complex expressions — is narrowly restricted by the properties of the initial state; these same restrictions underlie and account for the extraordinary richness and flexibility of the systems that emerge. It is a virtual truism that scope and limits are intimately related. The biological endowment that allows an embryo to become a mouse, with only the most meager environmental “information,” prevents it from becoming a fly or a monkey. The same must be true of human higher mental faculties, assuming that humans are part of the biological world, not angels.

We can think of the states attained by FL, including the stable states, as “languages”: in more technical terminology, we may call them “internalized languages” (I-languages). Having an I-language, a person is equipped to engage in the “creative use of language” that has traditionally been considered a primary indication of possession of mind; by Descartes and his followers, to cite the most famous case. The person can produce new expressions over an unbounded range, expressions that are appropriate to circumstances and situations but not caused by them, and that evoke thoughts in others that they might have expressed in similar ways. The nature of these abilities remains as obscure and puzzling to us as it was to the Cartesians, but with the shift of perspective to “internalist linguistics,” a great deal has been learned about the cognitive structures and operations that enter into these remarkable capacities.

Though the observation does not bear directly on the study of human language, it is nevertheless of interest that FL appears to be biologically isolated in critical respects, hence a species property in a stronger sense than just being a common human possession. To mention only the most obvious respect, an I-language is a system of discrete infinity, a generative process that yields an unbounded range of expressions, each with a definite sound and meaning. Systems of discrete infinity are rare in the biological world and unknown in non-human communication systems. When we look beyond the most elementary properties of human language, its apparently unique features become even more pronounced. In fundamental respects human language does not fall within the standard typologies of animal communication systems, and there is little reason to speculate that it evolved from them, or even that it should be regarded as having the “primary function” of communication (a rather obscure notion at best). Language can surely be used for communication, as can anything people do, but it is not unreasonable to adopt the traditional view that language is primarily an instrument for expression of thought, to others or to oneself; statistically speaking, use of language is overwhelmingly internal, as can easily be determined by introspection.

Viewed in the internalist perspective, the study of language is part of biology, taking its place alongside the study of the visual system, the “dance faculty” and navigational capacities of bees, the circulatory and digestive

systems, and other properties of organisms. Such systems can be studied at various levels. In the case of cognitive systems, these are sometimes called the “psychological” and “physiological” levels — again, terms of convenience only. A bee scientist may try to determine and characterize the computations carried out by the bee’s nervous system when it transmits or receives information about a distant flower, or when it finds its way back to the nest; that is the level of “psychological” analysis, in conventional terminology. Or one may try to find the neural basis for these computational capacities, a topic about which very little is known even for the simplest organisms; the level of “physiological” analysis. These are mutually supportive enterprises. What is learned at the “psychological level” commonly provides guidelines for the inquiry into neural mechanisms; and reciprocally, insights into neural mechanisms can inform the psychological inquiries that seek to reveal the properties of the organism in different terms.

In a similar way, the study of chemical reactions and properties, and of the structured entities postulated to account for them, provided guidelines for fundamental physics, and helped prepare the way for the eventual unification of the disciplines. 75 years ago, Bertrand Russell, who knew the sciences well, observed that “chemical laws cannot at present be reduced to physical laws.” His statement was correct, but as it turned out, misleading; they could not be reduced to physical laws in principle, as physics was then understood. Unification did come about a few years later, but only after the quantum theoretic revolution had provided a radically changed physics that could be unified with a virtually unchanged chemistry. That is by no means an unusual episode in the history of science. We have no idea what the outcome may be of today’s efforts to unify the psychological and physiological levels of scientific inquiry into cognitive capacities of organisms, human language included.

It is useful to bear in mind some important lessons of the recent unification of chemistry and physics, remembering that this is core hard science, dealing with the simplest and most elementary structures of the world, not studies at the outer reaches of understanding that deal with entities of extraordinary complexity. Prior to unification, it was common for leading scientists to regard the principles and postulated entities of chemistry as mere calculating devices, useful for predicting phenomena but lacking some mysterious property called “physical reality.” A century ago, atoms and molecules were regarded the same way by distinguished scientists. People believe in the molecular theory of gases only because they are familiar with the game of billiards, Poincare observed mockingly. Ludwig Boltzmann died in despair a century ago, feeling unable to convince his fellow-physicists of the physical reality of the atomic theory of which he was one of the founders. It is now understood that all of this was gross error. Boltzmann’s atoms, Kekule’s structured organic molecules, and other postulated entities were real in the only sense of the term we know: they had a crucial place in the best

explanations of phenomena that the human mind could contrive.

The lessons carry over to the study of cognitive capacities and structures: theories of insect navigation, or perception of rigid objects in motion, or I-language, and so on. One seeks the best explanations, looking forward to eventual unification with accounts that are formulated in different terms, but without foreknowledge of the form such unification might take, or even if it is a goal that can be achieved by human intelligence — after all, a specific biological system, not a universal instrument.

Within this “biolinguistic” perspective, the core problem is the study of particular I-languages, including the initial state from which they derive. A thesis that might be entertained is that this inquiry is privileged in that it is presupposed, if only tacitly, in every other approach to language: sociolinguistic, comparative, literary, etc. That seems reasonable, in fact almost inescapable; and a close examination of actual work will show, I think, that the thesis is adopted even when that is vociferously denied. At the very least it seems hard to deny a weaker thesis: that the study of linguistic capacities of persons should find a fundamental place in any serious investigation of other aspects of language and its use and functions. Just as human biology is a core part of anthropology, history, the arts, and in fact any aspect of human life, so the biolinguistic approach belongs to the social sciences and humanities as well as human biology.

Again adapting traditional terms to a new context, the theory of an I-language L is sometimes called its “grammar,” and the theory of the initial state S-0 of FL is called “universal grammar” (UG). The general study is often called “generative grammar” because a grammar is concerned with the ways in which L generates an infinite array of expressions. The experience relevant to the transition from S-0 to L is called “primary linguistic data” (PLD). A grammar G of the I-language L is said to satisfy the condition of “descriptive adequacy” to the extent that it is a true theory of L. UG is said to satisfy the condition of “explanatory adequacy” to the extent that it is a true theory of the initial state. The terminology was chosen to bring out the fact that UG can provide a deeper explanation of linguistic phenomena than G. G offers an account of the phenomena by describing the generative procedure that yields them; UG seeks to show how this generative procedure, hence the phenomena it yields, derive from PLD. We may think of S-0 as a mapping of PLD to L, and of UG as a theory of this operation; this idealized picture is sometimes said to constitute “the logical problem of language acquisition.” The study of language use investigates how the resources of I-language are employed to express thought, to talk about the world, to communicate information, to establish social relations, and so on. In principle, this study might seek to investigate the “creative aspect of language use,” but as noted, that topic seems shrouded in mystery, like much of the rest of the nature of action.

The biolinguistic turn of the 1950s resurrected many traditional

questions, but was able to approach them in new ways, with the help of intellectual tools that had not previously been available: in particular, a clear understanding of the nature of recursive processes, generative procedures that can characterize an infinity of objects (in this case, expressions of L) with finite means (the mechanisms of L). As soon as the inquiry was seriously undertaken, it was discovered that traditional grammars and dictionaries, no matter how rich and detailed, did not address central questions about linguistic expressions. They basically provide “hints” that can be used by someone equipped with FL and some of its states, but leave the nature of these systems unexamined. Very quickly, vast ranges of new phenomena were discovered, along with new problems, and sometimes at least partial answers.

It was recognized very soon that there is a serious tension between the search for descriptive and for explanatory adequacy. The former appears to lead to very intricate rule systems, varying among languages and among constructions of a particular language. But this cannot be correct, since each language is attained with a common FL on the basis of PLD providing little information about these rules and constructions.

The dilemma led to efforts to discover general properties of rule systems that can be extracted from particular grammars and attributed to UG, leaving a residue simple enough to be attainable on the basis of PLD. About 25 years ago, these efforts converged in the so-called “principles and parameters” (P&P) approach, which was a radical break from traditional ways of looking at language. The P&P approach dispenses with the rules and constructions that constituted the framework for traditional grammar, and were taken over, pretty much, in early generative grammar. The relative clauses of Hungarian and verb phrases of Japanese exist, but as taxonomic artifacts, rather like “terrestrial mammal” or “creature that flies.” The rules for forming them are decomposed into principles of UG that apply to a wide variety of traditional constructions. A particular language L is determined by fixing the values of a finite number of “parameters” of S-0: Do heads of phrases precede or follow their complements? Can certain categories be null (lacking phonetic realization)? Etc. The parameters must be simple enough for values to be set on the basis of restricted and easily obtained data. Language acquisition is the process of fixing these values. The parameters can be thought of as “atoms” of language, to borrow Mark Baker’s metaphor. Each human language is an arrangement of these atoms, determined by assigning values to the parameters. The fixed principles are available for constructing expressions however the atoms are arranged in a particular language. A major goal of research, then, is to discover something like a “periodic table” that will explain why only a very small fraction of imaginable linguistic systems appear to be instantiated, and attainable in the normal way.

Note that the P&P approach is a program, not a specific theory; it is a

framework for theory, which can be developed in various ways. It has proven to be a highly productive program, leading to an explosion of research into languages of a very broad typological range, and in far greater depth than before. A rich variety of previously-unknown phenomena have been unearthed, along with many new insights and provocative new problems. The program has also led to new and far-reaching studies of language acquisition and other areas of research. It is doubtful that there has ever been a period when so much has been learned about human language. Certainly the relevant fields look quite different than they did not very long ago.

The P&P approach, as noted, suggested a promising way to resolve the tension between the search for descriptive and explanatory adequacy; at least in principle, to some extent in practice. It became possible, really for the first time, to see at least the contours of what might be a genuine theory of language that might jointly satisfy the conditions of descriptive and explanatory adequacy. That makes it possible to entertain seriously further questions that arise within the biolinguistic approach, questions that had been raised much earlier in reflections on generative grammar, but left to the side: questions about how to proceed beyond explanatory adequacy.

It has long been understood that natural selection operates within a "channel" of possibilities established by natural law, and that the nature of an organism cannot truly be understood without an account of how the laws of nature enter into determining its structures, form, and properties. Classic studies of these questions were undertaken by D'Arcy Thompson and Alan Turing, who believed that these should ultimately become the central topics of the theory of evolution and of the development of organisms (morphogenesis). Similar questions arise in the study of cognitive systems, in particular FL. To the extent that they can be answered, we will have advanced beyond explanatory adequacy.

Inquiry into these topics has come to be called "the minimalist program." The study of UG seeks to determine what are the properties of language; its principles and parameters, if the P&P approach is on the right track. The minimalist program asks why language is based on these properties, not others. Specifically, we may seek to determine to what extent the properties of language can be derived from general properties of complex organisms and from the conditions that FL must satisfy to be usable at all: the "interface conditions" imposed by the systems with which FL interacts. Reformulating the traditional observation that language is a system of form and meaning, we observe that FL must at least satisfy interface conditions imposed by the sensorimotor systems (SM) and systems of thought and action, sometimes called "conceptual-intentional" (CI) systems. We can think of an I-language, to first approximation, as a system that links SM and CI by generating expressions that are "legible" by these systems, which exist independently of language. Since the states of FL are computational systems, the general properties that particularly concern us are

those of efficient computation. A very strong minimalist thesis would hold that FL is an optimal solution to the problem of linking SM and CI, in some natural sense of optimal computation.

Like the P&P approach that provides its natural setting, the minimalist program formulates questions, for which answers are to be sought — among them, the likely discovery that the questions were wrongly formulated and must be reconsidered. The program resembles earlier efforts to find the best theories of FL and its states, but poses questions of a different order, hard and intriguing ones: Could it be that FL and its states are themselves optimal, in some interesting sense? That would be an interesting and highly suggestive discovery, if true. In the past few years there has been extensive study of these topics from many different points of view, with some promising results, I think, and also many new problems and apparent paradoxes.

Insofar as the program succeeds, it will provide further evidence for the Galilean thesis that has inspired the modern sciences: the thesis that “nature is perfect,” and that the task of the scientist is to demonstrate this, whether studying the laws of motion, or the structure of snowflakes, or the form and growth of a flower, or the most complex system known to us, the human brain.

The past half century of the study of language has been rich and rewarding, and the prospects for moving forward seem exciting, not only within linguistics narrowly conceived but also in new directions, even including the long-standing hopes for unification of linguistics and the brain sciences, a tantalizing prospect, perhaps now at the horizon.

Noam Chomsky
Institute Professor at MIT

Preface

The present volume began as a synthesis of class notes for an introductory course in testing offered to graduate students of Teaching English as a Second/ Foreign Language. Chapters one through seven formed the point of departure for a one-semester course, supplemented with popular tests, articles on current testing techniques, and student projects in item writing and item and test analysis. To address the advent of important new developments in measurement theory and practice, the work was expanded to include introductory information on item response theory, item banking, computer adaptive testing, and program evaluation. These current developments form the basis of the later material in the book, chapters eight through ten, and round out the volume to be a more complete guide to language test development, evaluation, and research.

The text is designed to meet the needs of teachers and teachers-in-training who are preparing to develop tests, maintain testing programs, or conduct research in the field of language pedagogy. In addition, many of the ideas presented here will generalize to a wider audience and a greater variety of applications. The reader should realize that, while few assumptions are made about prior exposure to measurement theory, the book progresses rapidly. The novice is cautioned against beginning in the middle of the text without comprehension of material presented in the earlier chapters. Familiarity with the rudiments of statistical concepts such as correlation, regression, frequency distributions, and hypothesis testing will be useful in several chapters treating statistical concepts. A working knowledge of elementary algebra is essential. Some rather technical material is introduced in the book, but bear in mind that mastery of these concepts and techniques is not required to become an effective practitioner in the field. Let each reader concentrate on those individually challenging matters that will be useful to him or her in application. While basic principles in measurement theory are discussed, this is essentially a “how-to” book, with focus on practical application.

This volume will be helpful for students, practitioners, and researchers. The exercises at the end of each chapter are meant to reinforce the concepts and techniques presented in the text. Answers to these exercises at the back of the book provide additional support for students. A glossary of technical terms is also provided. Instructors using this text will probably want to supplement it with sample tests, publications on current issues in testing, and computer printouts from existing test analysis software. These supplementary materials, readily available, will enhance the concrete, practical foundation of this text.

Grant Henning

Table of Contents

Preface by Halliday	F10 F11
Preface by Chomsky	F14 F21 F24 F29
Preface	F29
CHAPTER 1. Language Measurement: Its Purposes, Its Types, Its Evaluation	1
1.1 Purposes of Language Tests	1
1.2 Types of Language Tests	4
1.3 Evaluation of Tests	9
1.4 Summary/Exercises	13
CHAPTER 2. Measurement Scales	15
2.1 What is a Measurement Scale?	15
2.2 What Are the Major Categories of Measurement Scales?	15
2.3 Questionnaires and Attitude Scales	21
2.4 Scale Transformations	26
2.5 Summary/Exercises	29
CHAPTER 3. Data Management in Measurement	31
3.1 Scoring	31
3.2 Coding	35
3.3 Arranging and Grouping of Test Data	36
3.4 Summary/Exercises	41
CHAPTER 4. Item Analysis: Item Uniqueness	43
4.1 Avoiding Problems at the Item-Writing Stage	43
4.2 Appropriate Distractor Selection	48
4.3 Item Difficulty	48
4.4 Item Discriminability	51
4.5 Item Variability	54
4.6 Distractor Tallies	55
4.7 Summary/Exercises	55

CHAPTER 5. Item and Test Relatedness	57
5.1 Pearson Product-Moment Correlation	57
5.2 The Phi Coefficient	67
5.3 Point Biserial and Biserial Correlation	68
5.4 Correction for Part-Whole Overlap	69
5.5 Correlation and Explanation	69
5.6 Summary/Exercises	70
CHAPTER 6. Language Test Reliability	73
6.1 What Is Reliability?	73
6.2 True Score and the Standard Error of Measurement	74
6.3 Threats to Test Reliability	75
6.4 Methods of Reliability Computation	80
6.5 Reliability and Test Length	85
6.6 Correction for Attenuation	85
6.7 Summary/Exercises	86
CHAPTER 7. Test Validity	89
7.1 What Is Validity of Measurement?	89
7.2 Validity in Relation to Reliability	89
7.3 Threats to Test Validity	91
7.4 Major Kinds of Validity	94
7.5 Summary/Exercises	105
CHAPTER 8. Latent Trait Measurement	107
8.1 Classical vs. Latent Trait Measurement Theory	107
8.2 Advantages Offered by Latent Trait Theory	108
8.3 Competing Latent Trait Models	116
8.4 Summary/Exercises	125
CHAPTER 9. Item Banking, Machine Construction of Tests, and Computer Adaptive Testing	127
9.1 Item Banking	127
9.2 Machine Construction of Tests	135
9.3 Computer Adaptive Testing	136
9.4 Summary/Exercises	140

CHAPTER 10. Evaluation of Language Instruction Programs

143

- 10.1 Affective Evaluation 143
- 10.2 Cognitive Evaluation 144
- 10.3 Evaluation of Delivery Systems 148
- 10.4 Statistical Procedures 152
- 10.5 Summary/Exercises 157

Bibliography 159

Appendix A: Tables 165

Appendix B: Answers to Questions 173

Appendix C: Glossary 189

Chapter One

Language Measurement: Its Purposes, Its Types, Its Evaluation

There could be no science as we know it without measurement. Testing, including all forms of language testing, is one form of measurement. Just as we weigh potatoes, examine the length of a piece of cloth, count eggs in a carton, or check the volume of a container of milk, so we test reading comprehension or spelling to determine to what degree these abilities are present in the learner. There is potential for error when we weigh potatoes. For example, the scale might not work properly, or it may not be highly sensitive, so that we must settle for a rough estimate of the correct weight. Furthermore, the potatoes might be wet or dirty, or there might be a few yams mixed in. In either case our measurement may be inaccurate.

In the same way, tests of language abilities may be inaccurate or *unreliable* in the sense that repeated measures may give different results. These measures may also be *invalid* in the sense that other abilities are mixed in. Our test of reading comprehension on closer examination may turn out to be a test of grammar or vocabulary, or at least a few such items may be “mixed in.” Tests, to be useful, must provide us with reliable and valid measurements for a variety of purposes.

1.1 Purposes of Language Tests

Diagnosis and Feedback

Perhaps the most common use of language tests, and educational tests in general, is to pinpoint strengths and weaknesses in the learned abilities of the student. We may discover through testing that a given student has excellent pronunciation and fluency of oral production in the language of interest, but that he or she has a low level of reading comprehension. On further testing, we might

find that a low or too highly specialized vocabulary is a major factor underlying low reading comprehension for this student. We might recommend suitable approaches for vocabulary expansion.

This use of tests, frequently termed *diagnostic testing*, is of value in that it provides critical information to the student, teacher, and administrator that should make the learning process more efficient. Without the specific information thus made available, the teacher might persist in teaching pronunciation to this student and fail entirely to address a weakness in the area of vocabulary.

Screening and Selection

Another important use of tests is to assist in the decision of who should be allowed to participate in a particular program of instruction. In every instructional program, teaching staff and facilities are limited in number and capacity. It becomes a matter of serious concern to find an equitable means of determining who should be allowed to participate when there are more applicants than spaces available. Such selection decisions are often made by determining who is most likely to benefit from instruction, to attain mastery of language or content area, or to become the most useful practitioner in the vocational domain represented.

Considerable controversy has arisen about the fairness of tests and the possibility that they may contain cultural or other biases against minority population groups when used for purposes of selection (Scheuneman, 1984). Some researchers seem to indicate that the effects of cultural bias, though present, may be small and actually in favor of minorities (Chen and Henning, 1985). However, most educators agree that some, though perhaps not entire, reliance must still be placed on test scores when screening or selection decisions are being made (Lennon, 1978). In order for such decisions to be fair, our tests must be accurate in the sense that they must provide information that is both reliable and valid.

In the area of language testing, a common screening instrument is termed an *aptitude test* (Carroll, 1965). It is used to predict the success or failure of students prospective in a language-learning program.

Placement

Closely related to the notions of diagnosis and selection is the concept of placement. In this case tests are used to identify a particular performance level of the student and to place him or her at an appropriate level of instruction. It follows that a given test may serve a variety of purposes; thus the *UCLA Placement Exam* may be used to assign students to levels as well as to screen students with extremely low English proficiency from participation in regular university instruction.

Program Evaluation

Another common use of tests, especially *achievement tests*, is to provide information about the effectiveness of programs of instruction. In this way the

focus of evaluation is not the individual student so much as the actual program of instruction. Therefore, group mean or average scores are of greater interest in this case than are isolated scores of individual students. Often one or more *pretests* are administered to assess gross levels of student proficiency or “entry behavior” prior to instruction. Following the sequence of instruction, one or more *posttests* are administered to measure postinstructional levels of proficiency or “exit behavior.” The differences between pretest and posttest scores for each student are referred to as *gain scores*.

Frequently in program evaluation tests or quizzes are administered at intervals throughout the course of instruction to measure “en route behavior.” If the results of these tests are used to modify the program to better suit the needs of the students, this process is termed *formative evaluation*. The final exam or posttest is administered as part of the process of what is called *summative evaluation* (Scriven, 1967).

Sometimes language programs may be evaluated by comparing mean posttest or gain scores of one program or partial program with those of other programs. Whatever the method of evaluation, the importance of sensitive, reliable, and valid tests is obvious.

Providing Research Criteria

Language test scores often provide a standard of judgment in a variety of other research contexts. Comparisons of methods and techniques of instruction, textbooks, or audiovisual aids usually entail reference to test scores. Even examination of the structure of language itself or the physiological and psychological processes of language use may involve some form of measurement or testing. If we are to learn more about effective methods of teaching, strategies of learning, presentation of material for learning, or description of language and linguistic processes, greater effort will need to be expended in the development of suitable language tests.

Assessment of Attitudes and Sociopsychological Differences

Research indicates that only from one-quarter to one-half of the variability in academic achievement is explainable in terms of cognitive aptitude (Khan, 1969). The importance of noncognitive factors in achievement is seldom more evident than in the field of language learning, where the level of persistence and application needed for significant achievement is enormous. Attitudes toward the target language, its people, and their culture have been identified as important affective correlates of good language learning (Naiman et al., 1978; Saadalla, 1979). It follows that appropriate measures are needed to determine the nature, direction, and intensity of attitudes related to language acquisition.

Apart from attitudes, other variables such as cognitive style of the learner (Witkin et al., 1977), socioeconomic status and locus of control of the learner (Morcos, 1979), linguistic situational context (Henning, 1978), and ego-permeability of the learner (Henning, 1979) have been found to relate to levels of language achievement and/or strategies of language use. Each of these factors in

A Guide to Language Testing Development · Evaluation · Research

Grant Henning

“This text appeals to me greatly because it is a paragon of clarity. Henning goes to great lengths to explain very technical concepts without relying on jargon. If computation is involved in a procedure, a step-by-step account is given with examples. Anyone who can read can learn about language test development from this text without a teacher . . . it is the best introductory testing book on the market!”

Kyle Perkins
Southern Illinois University

A systematic “how to” book covering an extensive range of practices and approaches, this comprehensive text introduces principles of test and questionnaire development for use in placement, achievement, evaluation, and research. It includes a thorough treatment of the estimation of reliability and validity, offering a variety of techniques to fit many measurement contexts. The inclusion of computational material and formulae offers specific information on testing development and evaluation, generally unavailable in a book of this kind.

Truly a practical guide to language testing, this volume will appeal to students, practitioners, and researchers who are preparing to develop tests, maintain testing programs, or conduct research in the field of language pedagogy.

