ED 213 035		CS 206 751
AUTHOR	Spandel, Vicki; Stiggins, Richard J.	
TITLE	Direct Measures of Writing Skill: Issues and Applications. Revised Edition.	
INSTITUTION	Northwest Regional Educational Laboratory, Portland, OR. Clearinghouse for Applied Performance Testing.	
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.	
PUB DATE	Aug 81	
GRANT	OB-NIE-G-78-0206	
NOTE	62p.; For related document see ED 196 038.	
EDRS PRICE	MF01/PC03 Plus Postage.	
DESCRIPTORS	*Educational Assessment; Educational Planning; Elementary Secondary Education; *Evaluation Methods; *National Surveys; Testing; Test Results; *Test Use; *Writing (Composition); *Writing Evaluation	

ABSTRACT

Intended for educators seeking information on direct writing assessments, this monograph describes general procedures for planning and conducting a writing assessment and strategies for tailoring that assessment to local needs. The introductory chapter offers a brief comparison of direct and indirect writing assessment methods, highlighting those features of direct assessment that make it the most popular approach. The status of writing assessment in American education is then summarized with emphasis on current patterns and developmental trends. The second chapter presents an overview of direct writing assessment procedures, touching on considerations in maximizing test quality, strategies for exercise development and alternative scoring approaches. The third chapter discusses selection of a writing assessment approach to suit a specific educational context such as program evaluation or student screening. Appended is a profile of statewide writing assessment, which provides information on the use of objective tests and or writing exercises, the developers of such exercises, kinds of writing assessed, kind of scoring method used, those using the results, and contact persons. (HOD)

Direct Measures of Writing Skill: Issues and Applications

REVISED EDITION

Vicki Spandel
Richard J. Stiggins

CAPT
Clearinghouse for Applied Performance Testing

NORTHWEST REGIONAL EDUCATIONAL LABORATORY
300 S.W. Sixth Avenue
Portland, Oregon 97204

January 1980
August 1981 (Revised Edition)

This work is published by the Clearinghouse for Applied Performance Testing (CAPT) of the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work contained herein has been developed under grant OB-NIE-G-78-0206 with the National Institute of Education (NIE), Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position of the National Institute of Education, and no official endorsement by the Institute should be inferred.

TABLE OF CONTENTS

PREFACE

The Clearinghouse for Applied Performance Testing (CAPT) has published a series of monographs on the assessment of writing proficiency. *Direct Measures of Writing Skill: Issues and Applications*, published in the first edition in January 1980, was the initial volume in that series. It presented perspectives on writing assessment as they had developed through the 1970s and included results of a 1979 national survey of statewide writing assessment programs compiled by Vicki Frederick of the Wisconsin State Department of Education.

This revised edition contains much of the same information found in the original. However, the views presented have been updated to reflect two additional years of writing assessment research and development Included in this edition are summarized results of a 1981 national survey of statewide and large-city school district writing assessment programs compiled by Michael McCready and Virginia Melton of Louisiana Technological University.

This monograph is written for educators interested in learning about procedures for the *direct* measurement of writing skills: that is, testing through the use of student writing samples. Minimum attention is given in this volume to the indirect assessment of writing skills via objective language usage tests. Material presented herein is directly useable by educators at all levels, from elementary, junior high and high school to postsecondary and state department levels.

Those interested in additional information on writing assessment are directed to three recent CAPT publications: *Using Writing Assessment in the Classroom: A Teacher's Handbook, A Directory of Writing Assessment Consultants* and *A Guide to Published Tests of Writing Proficiency*. The former provides teachers with strategies for using writing assessment methods to *teach* writing skills. The latter provides consumer information on available published tests of writing skill, and sources of technical assistance in developing and implementing writing assessment programs. Anyone interested in obtaining these publications is urged to contact CAPT for further details.

CAPT intends to continue its role in collecting, synthesizing and disseminating information on writing assessment. Readers are encouraged to submit comments and suggestions regarding this and other CAPT writing assessment publications.

Richard J. Stiggins
CAPT Coordinator

ACKNOWLEDGMENTS

The authors wish to thank CAPT staff members Nancy Bridgeford and Carol DeWitte for their valuable assistance in the production of this revised edition of *Direct Measures*. Nancy played a key editorial and production coordination role, and Carol provided excellent secretarial assistance.

Thanks are also due the NWREL Media Center for their production assistance and to the many people whose time and ideas contributed to development of the first edition.

Special thanks to Mike McCready and Virginia Melton of the Teacher Education Department, Louisiana Technological University, Ruston, Louisiana, for their willingness to share the results of their valuable research on state and local writing assessment programs.

Vicki Spandel
Richard J. Stiggins

CHAPTER I: Introduction to Writing Assessment

Until recently, those concerned with the large-scale assessment of writing proficiency relied predominantly on objective tests of language usage skill. Evidence of this fact can be found in the language skills tests included in standardized achievement batteries offered by publishers over the past 40 years, as well as in the language skills sections of the major national college entrance examinations. However, changing teacher attitudes and research and development efforts led by Educational Testing Service (ETS) and the National Assessment of Educational Progress (NAEP) have combined to shift the focus of writing assessment away from objective tests, toward the use of writing samples as the basis for judging proficiency. This new emphasis has been made possible in part through development of writing sample scoring procedures capable of producing valid and reliable results in an efficient, often cost effective manner.

The direct writing assessment techniques pioneered by ETS and NAEP are already being adopted by school districts, state education agencies, postsecondary institutions, and test publishers. Further, acknowledging the desirability of directly assessing writing proficiency, professional associations of English teachers are urging adoption of writing sample-based testing.

As a result of these developments, many educators are seeking information on direct writing assessment. This monograph has been prepared to help meet their needs. It offers the interested educator the basic information required to use educationally sound assessments of writing proficiency. It does not, however, present step-by-step instructions on how to measure writing skill. Those steps vary greatly from situation to situation, and should, whenever possible, be planned with the assistance of an experienced writing assessment consultant. The monograph does, however, describe general procedures for planning and conducting an assessment, and strategies for tailoring that assessment to local needs. Sources of additional information are also provided.

This introductory chapter offers a brief comparison of direct and indirect writing assessment methods, highlighting those features of direct assessment that make it the most popular approach. The status of writing assessment in American education is then summarized with emphasis on current patterns and developmental trends.

Chapter 2 presents an overview of direct writing assessment procedures, touching on considerations in maximizing test quality, strategies for exercise development and alternative scoring approaches. Chapter 3 discusses selection of a writing assessment approach to suit a specific educational context.

A Comparison of Direct and Indirect Writing Assessment

There are two viable approaches to the assessment of writing proficiency. One is the direct method. It relies on actual samples of student writing to judge writing proficiency. The second is the indirect method, which relies on objective tests. Research on the correlation between the two reveals a consistent and relatively strong relationship at various educational levels. Summarized here are six studies that correlated objective language usage test scores with scores obtained on writing sample-based assessments.

The results of these studies suggest that the two approaches assess at least some of the same performance factors; yet each deals with some unique aspects of writing skill. These similarities and differences relate to assessment focus, practical aspects of testing, characteristics of test exercises and aspects of test quality *

*For a more detailed discussion of these factors as they relate to direct and indirect assessment, see Stiggins (1981)

Researchers	Students Tested	N	Correlation
Godshalk, Swineford & Coffman (1966)	High school	646	46-.75
Breland, Colon & Rogosa (1976)	College	96	42
Breland & Gaynor (1979)	College	819	.63
		895	63
		517	58
Huntley, Schmeiser & Stiggins (1979)	College	50	.43-.67
Hogan & Mishler (1980)	Third graders	140	68
	Eighth graders	160	65
Moss, Cole & Khampalikit (1981)	Fourth graders	84	20-.68
	Seventh graders	45	.60- 67
	Tenth graders	98	72- 76

Assessment Focus. Direct and indirect writing assessments focus on different components of writing. Direct assessment measures actual composition skill. Indirect tests ability to use—or recognize proper use of—the conventions of effective writing: grammar, punctuation, sentence construction, organization, and so on. Direct assessment provides necessary and sufficient information for drawing conclusions regarding a student's writing proficiency. Indirect assessment, on the other hand, provides necessary—but not always sufficient—information for evaluating a student's writing proficiency.

An examination of traits measured in the two approaches reveals that indirect assessment *tends* to cover highly explicit constructs in which there are definite right and wrong responses (e.g., grammar is either correct or it is not). Direct assessment, on the other hand, *tends* to measure less tangible skills (e.g., persuasiveness), for which the concept of right and wrong is less relevant.

Practical Testing Considerations. Several important practical matters related to testing suggest additional differences between direct and indirect assessment.

For example, effective assessment requires appropriate attitudes on the part of test users. With direct assessment, users of the test results must be willing to invest the time, money and effort to conduct a writing assessment that calls for complex, often time consuming testing procedures. In the case of indirect assessment, users must be willing to accept a proxy measure: that is, a test that covers component skills of writing without actually requiring students to write. Given the appropriate attitudes, either direct or indirect assessment will most probably have its desired impact. If those attitudes are lacking, problems can be anticipated.

In either direct or indirect assessment, the examiner has two choices for test acquisition: (a) selecting an already existing test or (b) constructing a new test.

If one decides to use previously developed exercises and scoring criteria for direct assessment, then the following skills will be required of those conducting the assessment: (1) technical expertise in writing, to specify which writing skills will be assessed; (2) test evaluation skills to investigate available options and select test items that measure the skills to be assessed; and (3) organizational skills to set up, administer, score and report the results of the assessment.

Selecting an already developed objective test requires the expertise to determine the information needs of the test user and to review and select a valid and reliable test. In most cases, the user will also have to be skilled in interpreting and using norm-referenced standard scores.

Developing a new direct instrument, which involves creating a new set of exercises and criteria for scoring, also demands organizational skills and technical writing expertise. In addition, however, psychometric expertise is required in order to evaluate the validity and reliability of the assessment procedures, and refine exercises and criteria as necessary.

Developing a new indirect assessment or objective test instrument requires (1) technical expertise in writing to plan the assessment; (2) skill in item writing or selection; (3) organizational skills to pilot test, analyze and select the new items; and (4) psychometric expertise to evaluate the test's reliability and the validity.

In short, developing new instruments for either testing approach requires substantially more expertise and staff time than does using existing assessment instruments.

Characteristics of Test Exercises. There are some fundamental differences in the kinds of test exercises used in direct and indirect writing assessment First, the exercises differ in form Direct assessment exercises generally take the form of a short paragraph that invites the examinee to respond to a question, state an opinion, resolve an issue, explain a process, recount an event, or simply express his/her feelings. The exercise, if well constructed, identifies for the examinee the (1) form of writing to be produced, (2) audience to be addressed, and (3) purpose for the writing. Indirect assessment items frequently follow a multiple choice format, though fill-in questions are sometimes used. Various interlinear forms, as well as sentence combining items, are common.

As a result of differences in format, direct assessment exercises are considerably more flexible than indirect. With direct assessment, the stimulus can be auditory or visual and can be quite true to life (e.g., writing a job application letter). Indirect test items, on the other hand, are generally constrained by the multiple choice (or other) format. Therefore, while direct assessment exercises can be made to closely approximate "real world" writing, objective test items are somewhat more artificial.

Judging Test Quality. The factors commonly considered in judging the psychometric adequacy of a test are reliability and validity.

Reliability and validity considerations for direct and indirect measures are quite similar. In the case of direct measures, score stability is important over time, across exercises, across test forms and across raters. Consistency across raters is not an issue with indirect measures, however, since scoring is totally objective. In both cases, sources of inaccurate scores include poor test items and improper test administration. Sources of score inaccuracy unique to each approach include: (1) guessing on indirect measures, and (2) poor or inconsistent scoring of direct measures.

Validity considerations are similar. Content validity is relevant to both types of writing tests and should be verified in both via expert judgment. Criterion related validity, also important in both cases, can be verified through correlations with other indicators of writing proficiency.

Comparing Assessment Options. Direct and indirect approaches to writing assessment are perhaps best compared in terms of their relative advantages and disadvantages, and the primary ways in which each can be used.

The major advantages of the direct assessment option are

(1) the extent of information provided about examinees' writing proficiency, (2) potentially high fidelity (authenticity) of the exercise and response, (3) the adaptability of exercises to a variety of relevant real world writing circumstances, (4) high face validity, and (5) relatively low test development costs.

The major advantages associated with the indirect assessment are (1) high score reliability, (2) relatively low test scoring costs, and (3) high degree of control over the nature of the skills tested.

The disadvantages of the direct method include (1) high scoring costs, and (2) the potential lack of uniformity among examinees regarding the proficiencies assessed.

The disadvantages of the indirect method are (1) lack of fidelity to real world writing tasks, (2) heavy reliance on examinees' reading rather than writing proficiency, and in many cases (3) lack of face validity in the objective measure.

Writing assessment program developers would do well to keep these differences clearly in mind when planning an assessment program.

A Status Report on Writing Assessment Programs

In recent years, two national surveys of large-scale writing assessment programs have been conducted. The first was conducted by Frederick (1979) under the auspices of the Wisconsin Pupil Assessment Program of the Wisconsin Department of Public Instruction, and the second was conducted in 1981 by McCready and Melton of Louisiana Technological University with support from the National Institute of Education. Each survey, at the time it was conducted, provided very useful insights into the status of large-scale assessment, and taken together, the two surveys provide valuable perspectives regarding trends in writing assessment.

In 1979, 18 states were conducting writing assessment programs These assessments spanned the full range of grade levels. The typical assessment at that time covered about three grade levels, relied solely on a writing sample, or on a writing sample in combination with an objective test to judge proficiency, and involved holistic and/or primary trait scoring of the writing sample.

The status of writing assessment in 1981 is summarized in Table 1. Note that 24 states are currently conducting writing assessments relying predominantly on writing samples

Table 1
Evolution of Writing Assessments

	1979	1981	1981
	State Assessments	State Assessments	City Assessments
Conducting Assessments	18	24	20
Testing in Grades			
K	1		1
1	1	1	7
2	1	1	8
3	1	7	10
4	7	7	9
5	3	5	10
6	1	7	11
7	2	5	9
8	10	10	12
9	4	11	15
10	2	6	11
11	13	11	13
12	5	5	9
Mean Grades Tested Per State	2.7	3.2	6.3
Testing Strategy			
Objective only	1 6%	1 4%	3 5%
Writing Sample only	7 39%	12 50%	9 45%
Combination	10 55%	11 46%	8 40%
Scoring Method			
Holistic	6 33%	15 65%	8 50%
Analytical	1 6%	1 5%	5 31%
Primary Trait	6 33%	4 17%	0 0%
Combination	5 28%	3 13%	3 19%

scored holistically. Brief profiles are presented for statewide and large-city school district writing assessment programs. More detailed profiles are presented in the Appendix. The profiles are summarized in various ways in Table 2, which compares 1979 and 1981 assessments.

Several dimensions of this comparison are of interest. First, note that six more states have added assessment programs over the past two years. Note also that while statewide assessments in both 1979 and 1981 tended to begin testing in grades three or four, city schools tend to conduct a good deal

Table 2
Overview of Large-scale Assessment Programs

STATE	Grade(s) Tested	Type of Test	Writing Sample Scoring Procedure
Alabama	3, 6, 9	Writing Sample	Holistic
Delaware	1-8, 11	Objective Test Writing Sample	Primary Trait
California	3, 6, 12	Objective Test Writing Sample	Holistic
Florida	3, 5, 8, 11	Objective Test Writing Sample	Analytical
Hawaii	4, 8, 11	Writing Sample	Holistic
Idaho	9	Writing Sample	Holistic
Louisiana	3, 7, 10	Objective Test Writing Sample	Primary Trait
Maine	4, 8, 11	Writing Sample	Holistic
Maryland	9-12	Objective Test Writing Sample	Holistic
Massachusetts	7, 8, 9, 12	Writing Sample	Holistic Analytical
Michigan	4, 7, 10	Writing Sample	Primary Trait
Minnesota	4, 8, 11	Writing Sample	Primary Trait
Nevada	3, 6, 9-12	Objective Test (3, 6) Writing Sample	Holistic
New Hampshire	5, 9, 12	Writing Sample	Holistic
New Jersey	9	Objective Test Writing Sample	Holistic
New Mexico	10	Writing Sample	Holistic
North Carolina	11	Writing Sample	Holistic Analytical
Ohio	8, 12	Objective Test Writing Sample	Holistic

STATE	Grade(s) Tested	Type of Test	Writing Sample Scoring Procedure
Oregon	4, 7, 11	Objective Test Writing Sample	Holistic
Pennsylvania	5, 8, 11	Objective Test	
Rhode Island	4, 6, 8, 10	Objective Test Writing Sample	Holistic
South Carolina	6, 8, 11	Writing Sample	Holistic Analytical
Texas	3, 5, 9	Objective Test Writing Sample	Holistic
Wyoming	6, 9	Writing Sample	Holistic
CITY			
Little Rock, AR	1-11	Objective Test	
Phoenix, AZ	9-12	Objective Test Writing Sample	Analytical
Monterey, CA	1-12	Writing Sample	Holistic
Tallahassee, FL	1-8	Objective Test Writing Sample	Analytical
Atlanta, GA	1-12	Objective Test	
Des Moines, IA	9	Writing Sample	Holistic Analytical
Chicago, IL	9-12	Writing Sample	Analytical
Boston, MA	2, 5, 8	Writing Sample	Holistic
Wichita, KS	K-12	Writing Sample	Holistic
Baltimore, MD	1-9	Objective Test Writing Sample	Analytical
Detroit, MI	10-12	Objective Test Writing Sample	Holistic
Raleigh, NC	1-12	Objective Test Writing Sample	Teacher Option

STATE	Grade(s) Tested	Type of Test	Writing Sample Scoring Procedure
Albuquerque, NM	4, 6, 9-12	Objective Test (4, 6, 9) Writing Sample	Holistic
Santa Fe, NM	7-12	Writing Sample	Holistic Analytical
New York, NY	8, 11	Writing Sample	Holistic
Portland, OR	3-9	Objective Test	
Austin, TX	3, 9	Objective Test Writing Sample	Holistic
Madison, WI	5, 8, 11	Objective Test Writing Sample	Holistic Primary Trait
Seattle, WA	3, 6, 9-11	Writing Sample	Analytical
Laramie, WY	6, 9	Writing Sample	Holistic

of writing assessment as far down as grades one and two. Most assessment, however is conducted in junior and senior high school. The average number of grade levels tested is on the increase in statewide assessments. But neither the 1979 nor the 1981 averages on this variable compare to the city schools' average of 6.3 grade levels in each assessment.

With regard to writing assessment method, there is a relatively constant pattern over time and across settings. Large-scale assessments tend to rely on writing samples alone or writing samples in combination with objective tests. Sole reliance on objective tests is rare.

Procedures for rating writing performance have changed markedly over the past two years. In 1979, assessors tended to rely about equally on holistic and primary trait scoring. Little attention was paid to the analytical approach. In 1981, however, in both state and city programs, there has been a significant decline in the use of primary trait scoring, and a marked increase in the use of both holistic and analytical methods.

In sum, significantly more writing assessment is being conducted in 1981 than in 1979, and that assessment relies heavily on holistically scored samples of student writing as the basis for judging proficiency. For more detail on assessment programs consult the Appendix.

Still—No "Best" Answer

These are but a few of the many instances in which writing assessment is being successfully conducted on national, state, and local levels. The remainder of this monograph describes (1) some of the procedures used in various assessment contexts and (2) key measurement issues in the testing of writing skill.

The assessment of writing skill is a very complex task, because of the broad range of potentially relevant writing competencies and the difficulties in setting standards of acceptable performance. *There is not now, nor will there ever be, a single best way to assess writing skill. Each individual educational assessment and writing circumstance presents unique problems to the developer and user of writing tests. Therefore, great care must be taken in selecting the approach and the methods to be used in each writing assessment. Methods used in one context to measure one set of relevant writing skills should not be generalized to other writing contexts without very careful consideration of writing circumstances.*

Chapter II: An Overview of Direct Writing Assessment Procedures

The development and implementation of a high quality writing assessment program can be complex and expensive. This chapter outlines procedures for managing that complexity and ensuring sound assessment.

Ensuring High Quality Assessment

Two key considerations in determining the quality of writing assessment are the reliability and validity of the scores generated by the assessment. The exercise development and scoring procedures outlined in the following two sections of this chapter have been developed and refined specifically to ensure score reliability and validity. However, before describing those procedures, it may be useful to explain reliability and validity as they relate to direct writing assessment.

Reliability. To be useful for educational decisions, tests must yield scores that are consistent or reliable. When scores are unreliable, the assessment results can lead to erroneous conclusions or decisions. In writing assessment, score inconsistency can take any of several forms.

For example, suppose a writing assessment were administered to the same students twice, the second administration following a two- to three-week interval. And suppose that even though no writing instruction took place, the scores obtained the second time were totally different from those achieved the first time for nearly every examinee. The examiner would not know which score (if either) to depend on as the true reflection of the students' proficiency. Or suppose two writing exercises were developed to measure exactly the same skills and yet when both were administered to a student, the exercises resulted in totally different estimates of proficiency. Again, the examiner would not know which score was the better indicator of proficiency. Or, from a third perspective, suppose two judges read and evaluated a writing sample from the same student and drew totally different conclusions regarding the student's proficiency. In this case, as with the others, the examiner would not know which judgment to rely on. These three examples show how unreliability can manifest itself in the assessment of writing skill with writing samples.

When scores are unstable over time, differ across ostensibly equivalent writing exercises and or differ across independent evaluations of proficiency, there is reason to question the usefulness of the assessment procedures. However, when the procedures employed yield scores that are stable over time, across exercises and across independent evaluators, those scores can be confidently used for educational decisions. The test developer is responsible for (1) employing assessment development procedures that maximize score reliability, and (2) presenting systematic evidence of score reliability for review by users

Three factors are important in developing reliable tests. First, the writing skills to be measured must be clearly and concisely defined by writing experts. Only then is it possible to (1) demonstrate to users, exercise developers, and others precisely what skills are to be assessed; (2) judge exercise appropriateness; and (3) inform judges about the criteria for acceptable performance.

Second, there must be a clear and unambiguous link between the skills to be tested and the exercises developed. This interrelationship ensures that exercises give the competent writer the stimulus and opportunity to demonstrate whatever skill(s) the user wants to measure.

And third, judges must be carefully trained to conduct the evaluation according to prespecified criteria and agreed upon standards If these three guidelines are followed, chances are that scores will be consistent over time, across exercises, and across raters. If scores are found to be inconsistent, assessment procedures should be re-examined in light of these guidelines and revised accordingly.

Validity. Even if a developer of a direct writing assessment is successful in achieving score stability through careful skill identification, exercise development and evaluator training, the writing assessment developmental task is only partly completed. Attention must also be given to the validity of the assessment scores. The validity of a score depends on (1) the test used to generate that score, and (2) the intended purpose for that score. Intended purpose can be identified in a variety of ways, each of which can be considered a dimension of validity. Cronbach (1971) has identified a number of such dimensions that can be applied to the direct assessment of writing proficiency. For example, a test may be designed to measure a specific set of writing skills. If review of that test by qualified experts reveals that the exercises do indeed cover those skills, then the test is said to cover the intended content validly. It has achieved its content coverage purpose.

From a different but related perspective, a test that plays a significant role in educational decision making (e.g., provides a basis for placement or selection) should inspire confidence among users. The exercises must appear to assess truly important skills. If this face validity is missing, the test will not be used—regardless of the actual appropriateness of the exercises. It is important that the exercises seem appropriate even to the least sophisticated of the intended users.

There are other ways of revealing whether a test is achieving its intended purpose. For example, a test of writing proficiency is only one of many potential indicators of writing skill. If a test is valid, then scores should be consistent with (or reflect the same level of proficiency as) other indicators of writing skill: for example, performance on job-related or real-world writing tasks, amount of formal training in writing, grades received in writing courses, and/or scores achieved in other objective or writing sample-based tests of writing skill. To the extent that the writing assessment developer is able to show that performance on a newly developed writing assessment is consistent with performance on other writing-related tasks, the assessment has achieved its goal of reflecting writing proficiency.

Test purpose largely determines the requirements for documenting validity. For example, a direct writing assessment may be very general, or it may be narrowly focused to be precise and diagnostic. Suppose, for instance, that one wished to measure students' letter writing skills. A general exercise might present the student with these directions:

> Pretend that you are applying for a job as a salesperson with Acme, Inc. Write a letter to Acme explaining your interest and qualifications.

Because these instructions are very broad, responses can only be judged on general merit. Raters will likely consider such factors as word choice, sentence structure, organization, mechanics—in short, the kinds of things one would consider in judging any piece of writing. And the result will be a general profile of overall student writing performance. But suppose one wished to measure students' performance on **explicit** letter writing skills, in order to diagnose individual students' strengths and weaknesses. This would call for some modification in the item so that it might read as follows:

> Pretend that you are applying for a job as a salesperson with Acme, Inc Write a business letter addressed to Ms. Jones, Sales Manager of Acme, 2525 Main, Huntsville, New York 20201. Explain your interest and qualifications. Attempt to convince Acme that you're the best person for the job. Use proper business letter form.

These specific directions will allow responses to be judged according to explicit criteria: students' ability to be convincing and use proper business letter format. Responses to the first item could not be scored in this manner because the intended audience, purpose and expected letter format were not specified in the instructions. In summary, if diagnostic information is desired, items must be carefully structured to elicit the appropriate type of response. Evidence of success in achieving the desired level of precision should be included in validation research.

The purpose for testing may also be considered in terms of the specific educational decision in question. That is, a test may be intended to rank order examinees in terms of proficiency for selecting the most able for further training or the least able for remediation. Or the assessment may be intended to provide information for mastery/nonmastery decisions with regard to specific writing objectives. Because these are different purposes, the assessment strategies used to achieve them will differ. It is up to the developer to determine the usefulness and appropriateness of assessment procedures for meeting each specific decision-oriented purpose.

The essential point is that validity is a reflection of success in achieving the testing purpose. As with reliability, the test developer has two primary responsibilities: to maximize validity through careful test development and to report evidence of validity for users. Strategies for maximizing validity are similar to those for maximizing reliability. The writing skills to be assessed should be clearly and unambiguously defined. Both the skills and exercises developed to reflect those skills should carefully be reviewed by subject experts to ensure appropriateness. And once the test is administered and scored, scores should be related to other relevant writing proficiency indicators to be sure the assessment is focused on the desired dimensions of writing skill.

Developing Exercises

In the discussion that follows, a writing exercise is considered to comprise all stimulus materials and instructions used to define the writing task. Developing exercises for direct assessment of writing involves five carefully conducted steps. The first two steps are crucial for any writing assessment: (1) assessment planning and (2) exercise development. The remaining three steps, while very important, are not always implemented, depending on the resources available and the seriousness of the decisions to be made. These are (3) test specification and exercise review, (4) exercise pretesting and (5) final revision. Each of these five developmental steps is discussed in detail in the following paragraphs.

Assessment planning. The ultimate quality of any assessment is influenced more by the thoroughness and detail of its original blueprint than by any other factor. Several very important test design questions must be thoroughly considered. If each is not individually considered, the chances of creating a valid and reliable assessment—especially a writing assessment—are greatly reduced.

The first planning question concerns purpose. The sole reason for conducting any educational assessment is to provide information to facilitate some educational decision. Therefore, the primary step in writing assessment planning is to state precisely the specific educational decision to be influenced by the resulting scores. Potential decisions include (1) diagnosing individual student proficiency in specific writing skill areas; (2) rank ordering examinees with regard to general writing proficiency for selection or placement; and (3) assessing specific or general writing proficiency to evaluate the impact of an instructional program. (Additional decisions will be presented later.) Specific assessment strategies vary according to purpose. Therefore, the decision(s) to be facilitated must be clearly specified at the outset.

Second, test developers must determine the specific form of writing to be produced (e.g., essay, business letter, fiction), the audience to be addressed, and the purpose to be served in addressing that audience. Any given student's level of proficiency will vary as a function of writing form.

A third planning step calls for identifying the traits to be judged in evaluating writing skill and criteria or standards of acceptable performance for each trait selected. For example, organization, style, tone and sense of audience are typical *traits*: that is, elements of writing skill. In order to judge performance, however, evaluators need more than a list of traits. They need guidelines or *criteria* for determining good, poor or mediocre organization, style, and so on. The complexity of traits and criteria is a function of assessment purpose. A broad assessment of overall writing skill allows some flexibility in the specification of criteria. For a diagnostic assessment, on the other hand, both traits and scoring criteria must be delineated with great precision.

In summary, the writing assessment blueprint must include (1) the educational decision(s) to be facilitated, (2) the writing context (purpose, audience and type of writing to be required), and (3) the specific traits or skills to be judged along with criteria for evaluating performance.

Exercise development. Once planning is completed, the developmental goal becomes quite apparent: the design exercises that provide the competent student with the necessary stimulus and writing conditions to demonstrate his/her level of competency. In other words, the writing tasks must inform students of the purpose for the writing, the audience to be addressed and the type of writing expected (necessary conditions), while at the same time allowing students the latitude (e.g., sufficient exercises and time) to demonstrate their capabilities. It should be apparent that unless careful planning has preceded this step, appropriate exercise development will be difficult at best.

Here are some specific guidelines to b~ observed in constructing writing exercises: First, the exercise developer should recognize the impossibility of covering all possible instances of relevant writing. A realistic objective is to construct and include in the assessment an appropriate sample of relevant exercises. Based on student performance on that sample, one can generalize about expected performance in parallel contexts. To insure the appropriateness of these generalizations, however, samples must be carefully selected. For example, if one wishes to know whether students can write expository prosę for an academic audience, one exercise is probably not enough; two or three similar exercises may be necessary to ensure that the sample is sufficiently representative. At the same time, ability to construct other forms for other audiences—e.g., an entertaining piece of fiction for young children—is irrelevant to the testing purpose at hand.

To use another example, suppose the purpose of an assessment is to determine mastery of a single clearly focused writing objective: ability to present map directions effectively in written form. Enough examples of student performance should be gathered to ensure that addition of another exercise would not significantly alter any conclusions about student performance. In other words, exercises must be clearly focused and sufficient in number.

The reader may recognize that this issue of skill sampling is related to both reliability and validity, as described earlier. For example, it is important to provide enough samples of student writing to allow for stable scores (reliability), and to fairly and adequately sample the skill domain the test is intended to cover (validity).

Certainly the key question in all writing assessment is: How much writing is enough? There is no hard and fast answer. The number of exercises required and the length of those exercises are functions of the range of skills to be evaluated and the level of precision at which those skills are defined. Broader assessments covering many skills generally require more samples than precisely focused, narrow assessments. Recent research on this topic (Steele, 1979 and Breland, 1977) offers some guidance. The Steele research involved a broad assessment of end-of-college writing proficiency via three 20- to 30-minute writing exercises. Analysis of score consistency revealed that the use of only one or two exercises yielded unreliable scores. However, the use of all three exercises raised score consistency to an acceptable level. Further, the study revealed that the addition of more exercises beyond the original three would not significantly increase reliability. These results were supported by Breland's research which revealed that, in a similar college-level assessment, a single 20-minute exercise was incapable of yielding consistent scores.

Braddock, Lloyd-Jones and Schoer (1963) offer guidance from a different perspective as to the amount of writing needed to judge proficiency:

> Even if the investigator is primarily interested in nothing but grammar and mechanics, he should afford time for the writers to plan their central ideas, organization, and supporting details, otherwise their sentence structure and mechanics will be produced under artificial circumstances Furthermore, the writers ordinarily should have time to edit and proofread their work after they have come to the end of their papers. . . *Investigators should consider permitting primary grade children to take as much as 20 to 30 minutes, intermediate graders as much as 35 to 50 minutes, junior high school students 50 to 75 minutes, high school students 70 to 90 minutes, and college students two hours (to demonstrate proficiency).* [Emphasis added.]

Exercises should frame a clear and concise writing task so that students fully understand what is required—whether or not they can fulfill the requirements. Time pressure is undesirable, it is an artificial imposition that may not replicate the circumstances in which real life writing occurs. Items should offer the writer a realistic, sensible challenge so as to maintain interest. Varied stimulus materials (written, auditory, or visual) should be used. Most important, examinees must be given time to think, organize, write, reread and revise.

Some writing assessments have attached great importance to revision. As Rivas (1977) notes:

> Rewriting skills are often considered to be the essence of good writing All of us can express ourselves in some form, however ambiguous or inappropriate, but a good writer knows how to revise such preliminary statements so that they become less ambiguous and more appropriate.

Part of NAEP's 1974 writing assessment called for writing and rewriting the same copy in an attempt to get at revision (Rivas, 1977). Students were asked to write a class report about the moon, given certain facts. They were given 15 minutes to write the first draft, using a pencil. Upon finishing, they were given 13 minutes to revise the first draft, using a blue pen so that any changes would stand out clearly. They were told to make any changes they wished, including crossing out words or rewriting if necessary; rewriting was not required, however. Papers were scored for overall organization (based on the quality of the revision), and were categorized to indicate the kinds of revisions attempted: cosmetic (improved legibility), mechanical, grammatical, transitional, informational, holistic (complete rewriting), and so on. Though some educators might feel the test was not a true measure of revision skills (many students, for reasons unknown, attempted no revision), the NAEP moon test represents at least a step toward development of a proper revision test.

Clearly, attention must be given to editing and revision as part of any writing assessment, whether by providing sufficient time and opportunity for the examinees to revise on their own, or by providing specific instructions to revise, as NAEP did. If extensive revision (beyond proofreading for spelling and other mechanical errors) is desired, it will be necessary to construct the assessment to allow students time for proper reflection—just as in a real-life writing situation. It will not be sufficient merely to give students an additional five or ten minutes at the end of a writing exercise to "fix things up." A better approach might be to allow students time to write one day, time to revise on a subsequent day. This kind of provision may increase administration time and costs. However, it will also provide a more relevant (i.e., true to real life) test of revision skills than one-session assessment.

Review of specifications and exercises. Whenever possible, the writing and assessment personnel responsible for assessment specifications and writing exercises should present their work to an independent group of writing and measurement specialists for review and formative evaluation. This review should cover—

1. The purpose for the assessment (decision to be made).

2. The definition of the assessment context (form of writing, audience and reason for writing).
3. The criteria (skills to be assessed) and standards of acceptable performance.
4. Relevance of exercises in terms of skills to be assessed.
5. Representativeness of exercises in terms of the domain of possible exercises.
6. Sufficiency of the exercises in providing students with the opportunity, in terms of time and tasks, to demonstrate proficiency.
7. Clarity and conciseness of prescribed writing tasks.
8. Level of interest and challenge conveyed in stimulus materials and writing instruction.
9. Adequacy of instructions and opportunity for revision, if that is a desired part of the assessment.

As the importance of an educational decision and/or as the number of students to be included in the writing assessment increases, the importance of independent review increases also. Thus, review is less critical with small-scale, local or classroom assessments than with large-scale assessments on which selection decisions are often based.

Exercise pretesting. Whenever possible, exercises should be administered to a sample of students prior to actual full-scale administration so that potential problems can be identified and corrected. Pretesting procedures should closely approximate actual administration in terms of type (though not number) of pretest students, conditions (e.g., facilities, time limits, methods for providing directions) and scoring procedures. Developers should then independently evaluate results, attending to (1) the level of proficiency demonstrated (and whether that level seems to fluctuate from exercise to exercise), (2) the nature of the responses produced (in terms of quality, appropriateness, length and enthusiasm) (3) the consistency of ratings across independent evaluations, and (4) the apparent clarity of instructions to students. Exercises that appear to yield inconsistent or repeatedly low quality results can be identified and the reasons for apparent problems discussed. Often, exercises can be adjusted. As with independent exercise review, the importance of pretesting increases with the scope and importance of the assessment.

Final exercise revision. The final step in exercise development is to revise exercises on the basis of the review and pretest results. As final revisions are made, developers should continue to ensure reliability and validity of scores through careful use of test specifications, exercise development and preparation for scoring.

Procedures for Scoring Writing Samples

Many forms of objective tests can be machine scored. Writing tests that rely on writing samples, however, require individual hand scoring by qualified persons trained to apply agreed upon criteria and performance standards. Several different methods have been devised for scoring writing samples depending on the assessment purpose. The most appropriate method in any given situation depends upon what information one wishes to gain through scoring, how that information will be used, and what resources are available. Some scoring methods are more complicated—and therefore more costly—than others. The purpose of this section is to present a comparative overview of the general advantages and disadvantages inherent in each of five approaches: holistic scoring, analytical scoring, primary trait scoring, scoring for mechanics and grammar, and T-unit analysis.

Holistic scoring. In holistic scoring, raters review a paper for an overall or "whole" impression. Specific factors such as grammar, usage, style, tone and vocabulary undoubtedly affect the rater's response, but none of these considerations is directly addressed. As with all rating methods, raters must be carefully trained to conduct the evaluation. The purpose of training is to minimize (at least temporarily) the effects of individual biases by helping raters internalize an agreed upon set of scoring standards. It is generally recommended that raters be experienced in language arts, familiar with pertinent terminology and practiced in rating student papers at the level for which they will be scoring. Consistency—both among raters and among scores assigned by a single rater—is very important in holistic scoring. Initial training takes about half a day, but it is also necessary to build in time for "refresher" sessions throughout the course of any scoring activity.

Papers are rated on a numerical scale. NAEP has used both 4-point and 8-point scales. Four-point scales are most common. An even-numbered scale is recommended because it eliminates the convenience of a mid-point "dumping ground" for borderline papers.

Prior to actual scoring, the trainer and the most qualified or experienced raters review a subset of the papers to be scored in order to identify "range finders." These are papers that are representative of all the papers at a given scoring level. With a four-point scale, for example, there would be range finders for the 4, 3, 2 and 1 levels. Range finder papers must be so typical of papers at a given level that virtually all readers agree on the assigned score. This is vital because range finders are used in training, and later used as models to assist raters during scoring. Trainers and their assistants may have to read dozens of papers in order to find the "typical" range finder papers with which everyone is satisfied. For training purposes, it is advisable to have at least two (preferably more) range finders at each level.

Trainers do not work from any predetermined set of criteria in identifying range finders. They may, of course, discuss their findings and observations during the process. But it is important to realize that in holistic scoring, there is no preconceived notion of the "ideal" paper. A paper assigned a score of 4 will simply be a relatively high quality paper within a given group, it may or may not be an excellent paper in its own right. As Brown (1977) notes, "It is possible that all of the papers at the top of the score are horribly written. They may be better than the rest, but still may be unacceptable to most teachers of composition." If one has in mind some specific criterion of performance that students must meet, holistic scoring will not be appropriate. Scoring levels are set from within, irrespective of external standards.

Despite personal preferences, the holistic approach quickly produces marked consistency among raters—in virtually any group. This may be partly the result of peer pressure. But more likely it suggests that language arts people can agree—though the bases for their conclusions may differ—on what constitutes a relatively good and a relatively poor paper. Interrater reliability (that is, agreement between any two raters) can be expected to run from about .60 to .80 (Diederich, 1974). It may be higher in a few cases, depending upon the background of the raters and the amount of training time allowed (so that raters can internalize the system).

All papers should be read by at least two raters to minimize the chance of error resulting from rater fatigue, prejudice or other extraneous factors. ACT has achieved an interrater reliability of .75 using two raters and three writing samples (ACT, 1979). Increasing the number of raters beyond two does not seem to enhance score reliability (Steele, 1979).

Scores may be added or averaged across raters to determine a final score. Disagreements of more than one rating point should be resolved by a third reader or through discussion by the disagreeing raters. Such disagreements can typically be expected to occur in fewer than 5 percent of all cases if careful assessment planning and rater training is conducted.

Holistic scoring is rapid and efficient. Depending on the length of student responses, experienced raters can usually go through 30 to 40 papers per hour (though inexperienced raters cannot be expected to match this rate). Six hours of scoring per day is considered about maximum to maintain high reliability. Scoring is intensive work; short hours with frequent break periods yield the best results.

Because scoring levels are never defined, holistic scoring does not permit the reporting of specifics on student performance. After reading hundreds of papers, however, raters typically have a supremely clear notion of what factors influenced them to assign particular scores. For reporting purposes they may translate those observations into level definitions. Suppose, for example, that students were asked to write a job application letter. One might then say that a "typical" 4 paper used proper business letter format, used vocabulary and tone appropriate to the occasion, described the student's qualifications in a way that reflected a clear understanding of job requirements (as presented in the item), and reflected consistently good sentence structure, correct mechanics, and so on. Such a definition would not necessarily apply in total to every 4 paper, but would certainly capture the essence of papers at that level and help make results meaningful to parents and other audiences. Presentation of such definitions in conjunction with sample student papers can be an extremely effective reporting technique.

Analytical scoring. Analytical scoring involves isolating one or more characteristics of writing and scoring them individually. Analytical scoring is most appropriate if one wants to measure (and report) students' ability to deal with one or more specific conventions of writing: punctuation, organization, syntax, usage, creativity, sense of audience, and so on. Traits must be explicit and well defined so that all raters understand and agree upon the basis for making judgments. In addition, it is necessary to delineate in advance specific and complete criteria for judging each trait. In analytical scoring, raters rely on written guidelines—not range finders—to assist them in assigning scores. Ideally, raters should have a chance to participate in selecting traits and establishing criteria. This promotes understanding of and agreement with criteria, and ultimately enhances interrater reliability. Except for the setting of criteria, training and administration procedures are similar to those for holistic scoring.

Analytical scoring provides data on specific aspects of student writing performance. But does it really reveal whether, in general, students write well? The answer depends on (1) whether enough traits are analyzed to provide a comprehensive picture, and (2) whether those traits analyzed are significant—that is, whether they actually contribute to good writing. In an effort to identify those characteristics that seem most to influence a reader's judgment about the quality of a piece of writing, Diederich (1974) performed a content analysis on a sample of student essays scored holistically. Marginal comments were invited (as would not be the case in a traditional holistic session), and later tallied to isolate those factors that seemed to influence experienced raters' scores most. Here, in order of significance, are the factors Diederich isolated through that study:

1. Ideas
2. Mechanics (including usage, punctuation and spelling)
3. Organization
4. Wording
5. Flavor (or style)

Of course, individual examiners may identify other traits they wish to score. However, this list of traits permits a reasonably comprehensive analysis of writing.

Factor-by-factor analysis of writing elements is more time consuming than holistic scoring. Depending on how many factors one looks at, it requires two to three times as long (or more) to rate a paper analytically as it does holistically.

Analytical rating has been criticized because there is some indication it produces a "halo" effect; that is, students who are rated high on one trait will tend to be rated high on all traits. Page (1968) explains,

> A constant danger in multi-trait ratings is that they may reflect little more than some general halo effect, and that the presumed differential traits will really not be meaningful. . . . We find (in our research) a very large halo, or tendency for ratings to agree with each other.

Despite these disadvantages, however, analytical scoring has one great advantage: it provides potential for trait-by-trait analysis of students' writing proficiency.

Primary trait scoring. Primary trait scoring is similar to analytical scoring in that it focuses on a specific characteristic (or characteristics) of a given piece of writing. However, while analytical scoring attempts to isolate those characteristics important to any piece of writing in any situation, primary trait analysis is rhetorically and situationally specific. The most important—or **primary**—trait(s) in a letter to the editor will not likely be the same as that (those) in a set of directions f(, assembling a bicycle.

The primary trait system is based on the premise that all writing is done in terms of an audience, and that successful writing will have the desired effect upon that audience. For example, a good mystery story will excite and entertain the reader; a good letter of application will get the interview. In a scoring situation, of course, papers must be judged on the **likelihood** of their producing the desired response.

Because they are situation-specific, primary traits differ from item to item, depending on the nature of the assignment. Suppose a student were asked to give directions for driving from his/her home to school. The primary trait might then be sequential organization, for any clear, unambiguous set of directions would necessarily be well organized with details presented in proper order. As Mullis (1974) points out, "Successful papers will have that [primary] trait; unsuccessful papers will not—regardless of how well written they may be in other respects."

Raters determine that some traits are essential to success in a given assignment. However, additional traits that contribute but are not necessarily essential to the success of a paper are termed "secondary" traits and may also be included in the evaluation, if they can be clearly defined and exemplified for raters Scores may be weighted to show the relative importance of various traits, if desired, then totalled to indicate the overall quality of the paper.

The first step in primary trait scoring is to determine which trait or traits will be scored. The second is to develop a scoring guide to aid raters in assigning scores. To illustrate, consider the following guide developed by NAEP for scoring "letters to the principal on solving a problem in school." It was determined that a good letter would identify the problem, present a solution, and explain how that solution would improve the school Here are NAEP's criterion levels:

1. Respondents do not identify a problem or give no evidence that the problem can be solved or is worth solving.
2. Respondents identify a problem and either tell how to solve it *or* tell how the school would be improved if it were solved.
3. Respondents identify a problem, explain how to solve the problem, *and* tell how the school would be improved if the problem were solved.
4. Respondents include the elements of a "3" paper. In addition, the elements are expanded and presented in a systematic structure that reflects the steps necessary to solve the problem (Mullis, 1974).

Range finder papers may be used in addition to the scoring guide. This practice is not common, however, for many raters find it cumbersome to rely on two points of reference.

All raters should be familiar with the rationale underlying the primary trait system, and with the level definitions to be used in scoring. Raters must accept the fact that they will be looking for specific, well-defined traits, and be cautious about allowing extraneous criteria to influence scoring NAEP recommends that raters prescore (for practice) at **least** 10 sample papers at each level during training in order to become comfortable with applying the criteria (Mullis, 1974).

As with analytical scoring, defining criterion levels is the most time consuming step. It may be necessary to "test" numerous definitions on sample papers in order to come up with a set that works. Herein lies a strong argument for keeping the list of traits to be scored brief. On an average, count on a day of trial and error, discussion and debate for **each trait** to be defined. This may sound time consuming, but the quality and clarity of the final definitions, and the ease with which they can be applied, will readily justify the time spent.

Like analytical scoring, primary trait scoring can allow the reporting of student performance with respect to specific characteristics: e.g., organization, awareness of audience. For this reason, primary trait scoring is greatly favored over holistic scoring in contexts where more precise information is needed. But this advantage should be carefully weighed against the time and effort required to set up a workable primary trait scoring system. Aside from adopting already written criteria (e.g., from NAEP), there are no known shortcuts.

Scoring language usage and mechanics. Of the types of scoring mentioned thus far, the scoring of writing mechanics is the most time consuming, and the most complex approach for which to provide training. This realization often comes as a great surprise to inexperienced raters, who may look on mechanics as a rather cut and dried affair—until faced with the prospect of setting up a scoring system.

The fact is, the standards of appropriate usage are subject to continual change through popular usage. So rapid has that change become now that even usage textbooks sometimes reflect different notions of what is appropriate. For the sake of consistency in scoring mechanics, it is necessary that a fairly comprehensive guide be developed. It is possible, of course, to use a standard reference—an English handbook—for this purpose. But raters must agree to abide by the document, and if there are too many areas of disagreement, it may be simpler to design their own. Whatever the decision, it is imperative that everyone agree to score according to the rules of the guide, **regardless of personal preference**. Otherwise, the inconsistency will render the scores useless.

Several other decisions must be made as well:

1. Whether to count errors of commission and errors of omission equally.

2. Whether to require formal usage, or to base guide rules on informal usage.

3. Whether to count errors involving concepts or rules with which students may not be familiar (e.g., seventh graders may not have been taught proper use of colon and semicolon—should this be considered?).

4. Whether to count every identifiable error or to focus on specific areas for easier reporting of results.

In addition, raters must establish a workable rating scale. If they choose to retain a 4-point scale, for example, it will be necessary to determine how many errors will be allowed in a 4 paper, how many in a 3 and so on.

One additional step necessary in scoring writing mechanics is obtaining an accurate word count for each paper. Errors can then be tabulated per 100 words. Analyzing errors in this way does not penalize those who write long responses, or give unfair advantage to those who write very little.

Test administrators should be cautioned about scoring mechanics as one trait within a primary trait system. As the foregoing discussion indicates, it is far more time consuming to score than other traits, and demands a number of special considerations. Therefore, test administrators should weigh carefully the advantages and disadvantages of such a combined approach.

Educators considering using the direct assessment approach to evaluate mechanics should remember that understanding of such usage elements as punctuation, grammar, diction, and sentence structure can be very efficiently, validly and reliably assessed using available indirect assessment measures. For mechanics or usage assessment, very careful consideration should be given to the objective test because it forces examinees to demonstrate explicit ability to deal effectively with the precise elements being tested. If a writing sample is used to assess these elements, examinees will typically avoid language constructions which they are unable to use effectively. Further, inconsistencies in usage patterns will make comparisons among examinees, on the basis of mechanics, difficult if not impossible. Such comparisons are generally possible with objective usage tests. In addition, because a writing sample taps but a small, arbitrary portion of an examinee's proficiency in writing mechanics, results cannot appropriately be used in diagnosis, whereas objective test results may be quite suitable for this purpose.

T-unit analysis. The concept of T-unit analysis was introduced in the 60s, and has gained popularity ever since as a means of measuring writing sophistication. A T-unit may be thought of as an independent clause plus whatever subordinate clauses or phrases accompany it. In simple terms, a T-unit is the smallest group of words in a piece of writing that **could** be punctuated as a sentence (T stands for "terminable"). Consider the following passage:

> I yelled at my cat Manfred and he ran away, but he came home when he got hungry.

This passage has only one terminal mark of punctuation as written, but actually contains three T-units:

- I yelled at my cat Manfred
- and he ran away,
- but he came home when he got hungry.

Each of these T-units is an independent clause that **could** be punctuated as a sentence. Note that T-unit analysis is **independent of punctuation**; a writer may or may not punctuate T-units as sentences.

Studies have shown that T-unit length tends to increase with the age and skill of the writer* (Hunt, 1977). In addition, it has been demonstrated that with increased skill, writers can incorporate a greater number of distinct concepts into a single T-unit. Consider the following example, using six short sentences, each of which consists of one T-unit, abstracted from a longer piece:

1. Aluminum is a metal.
2. It is abundant.
3. It has many uses.
4. It comes from bauxite.

*There are notable exceptions, therefore, this tendency cannot be applied as a general rule Highly experienced, sophisticated writers may consistently use short T-units Conversely, the use of lengthy T-units does not of itself render one a skillful writer.

Table 3

A Comparison of Scoring Methods for Direct Writing Assessment

DESCRIPTOR	HOLISTIC	ANALYTICAL
GENERAL CAPABILITIES	Comprehensive, general picture of student performance, writing viewed as a unified coherent whole. Applicable to any writing task	Thorough, trait by trait analysis of writing, provides comprehensive picture of performance *if* enough traits are analyzed, traits are those important to *any* piece of writing in *any* situation (e g , organization, wording, mechanics)
RELIABILITY	High reliability if standards are carefully established and raters are carefully trained	High reliability if criteria and standards are well defined, and careful training is conducted
PREPARATION TIME	Up to one day per item to identify range finder (model) papers, up to one-half day to train readers using 4-point scale, full day to train with 8-point scale	One full day to identify traits, one day per trait to develop scoring criteria (unless traits and criteria are borrowed from another source), one to two days to review results of pilot test and refine traits or criteria as necessary, one-half day to train raters
READERS	Qualified language arts personnel recommended, high reliability can be achieved with non-language arts readers given sufficient training	Qualified language arts personnel recommended
SCORING TIME	One to two minutes per paper (experienced readers may read faster)	One to two minutes per paper per trait
CLASSROOM USE	May be adapted for use in class	May be adapted for use in class
REPORTING	Allows reporting on students' overall writing skill	Allows reporting of student performance on wide range of generalizable traits (i e , the qualities considered important to all good writing)
GROUP/ SAMPLE SIZE*	Primarily usable with a larger sample, with a small sample, responses may be difficult to scale	Best with smaller samples, extensive scoring time *may* make costs prohibitive with larger groups

*These are very general guidelines Due to the nature of the scoring-cost/amount-of-information trade-off across scoring methods, readers are urged to seek the technical assistance of a qualified writing assessment specialist if there is a question regarding the appropriate use of available scoring resources

PRIMARY TRAIT	WRITING MECHANICS	T-UNIT ANALYSIS
Highly focused analysis of situation-specific primary trait (and possibly secondary traits). provides specific information on a narrowly defined writing task (e g. ability to recount details in chronological order)	Can provide either a general or a specific profile of the student's ability to use mechanics properly.	Provides a measure of syntactical sophistication
High reliability if criteria and standards are well defined. and careful training is conducted.	High reliability if given sufficient training time and authoritative. complete, acceptable guidelines (e g . an English handbook).	High reliability provided trained and experienced raters are used
One full day to identify traits, one day per trait to develop scoring criteria (unless traits and criteria are borrowed from another source); one to two days to review results of pilot test and refine traits or criteria as necessary. one-half day to train raters	One to two days to set up a scoring system (unless borrowed from another source). Minimum of one day to internalize the scoring system and practice scoring	Half day to full day. depending on raters' previous experience
Qualified language arts personnel recommended. non-language arts staff may be able to score some traits.	Qualified language arts personnel recommended.	Raters *must* be experienced language arts personnel. preferably those already familiar with the concept of T-unit analysis
One to two minutes per paper per trait	Five minutes or more per paper. depending on number of criteria	Varies greatly, depending on raters' skill.
May be adapted for use in class	May be adapted for use in class	May be adapted for use in class
Allows reporting of student performance on one or more situation-specific traits important to a particular task	Allows reporting of group or individual data on students' general strengths or weaknesses in mechanics	Allows group or individual reporting on syntactical sophistication
Generally more cost-effective with smaller samples. depending on the number of traits to be scored (with one trait, sample size is not an issue)	Best with smaller samples. extensive scoring time *may* make costs prohibitive with larger groups	Best with smaller samples. extensive scoring time *may* make costs prohibitive with larger groups

5. Bauxite is an ore.

6. Bauxite looks like clay.

Here's how a fourth grader rewrote the passage:

> Aluminum is a metal and it is abundant. It has many uses and it comes from bauxite. Bauxite is an ore and looks like clay. (6 sentences to 5 T-units)

The revision of a typical eighth grader:

> Aluminum is an abundant metal, has many uses, and comes from bauxite. Bauxite is an ore that looks like clay. (6 sentences into 2 T-units)

And finally, the revision of a skilled adult, a professional writer:

> Aluminum, an abundant metal with many uses, comes from bauxite, a claylike ore (6 sentences into 1 T-unit)

T-unit analysis and review of conversions (from simple sentences into T-units) provide a good measure of sentence maturity and of a student's ability to consolidate multiple thoughts.

Sophisticated, condensed writing has undeniable appeal. T-unit analysis used in conjunction with holistic scoring is likely to reveal that the highest scored papers (i.e., those that appealed most to readers) were in fact those with the most sophisticated use of T-units.

T-unit analysis is still in the experimental stages It is time consuming and costly to conduct. Moreover, it can [illegible]e done by highly trained language arts specialists. Furt[illegible] research and use may, however, reveal more widespread applicability than has so far been anticipated. Two interesting footnotes: syntactical maturity is apparently reflected in oral speech as well as in writing, and such maturity can be enhanced through a sentence combining curriculum (Hunt, 1977).

A Comparison of Scoring Methods

Table 3 offers a comparative ov rview of the scoring procedures discussed in this section, focusing on several key descriptors.

Chapter III: Adapting Writing Assessment to Specific Purposes

Educational tests have only one function: to facilitate educational decision making. A test should not be administered, therefore, until the decision or decisions that rest on the results of that test have been clearly articulated. This applies to all tests, including writing tests.

In many educational contexts, writing tests can be and are being used effectively. For example, tests can play a role in instructional management decisions. Such decisions include (1) the diagnosis of individual learner strengths and weaknesses for instructional planning, (2) the placement of students into the next most appropriate level of instruction, and (3) educational and vocational planning as part of student guidance and counseling.

Tests can also be administered at key points in an educational program to check student development in order to (1) screen the admission to an advanced or remedial program, or (2) certify minimum proficiency (e.g., for high school graduation).

And finally, tests can be used for program evaluation purposes such as in (1) large-scale survey assessment, (2) formative program evaluation, and (3) summative program evaluation.

In the discussion that follows, each of these eight contexts is described in terms of the decision to be made, the primary decision makers, and the type of writing skill information needed to make the decision. Decision makers include students, parents, teachers, administrators (including specific project or program administrators, as well as building-, district- and state-level administrators), guidance counselors, and the public (including taxpayers and elected officials).

Using Tests to Manage Instruction

Diagnosis. Teachers often use tests and other performance indicators to track each student's level of development, thereby determining where that student is in the instructional sequence, and anticipating the next appropriate level of instruction. Diagnostic data gathered via direct writing assessment can help individualize instruction by simplifying student grouping or instructional scheduling decisions. In addition, diagnostic writing skill data gathered over time may provide a basis for grading or communicating progress to parents.

Placement. Decision makers such as teachers and educational administrators must place each student at the level of instruction best suited to his/her skills. Typically, they use such performance indicators as writing skill tests, previous courses completed, and grades to rank order students along a continuum of writing skill development, then place them in the appropriate course.

Guidance and Counseling. In deciding their future educational or vocational activities, students need to know how their writing skill compares to that of other students with whom they could compete. Performance indicators like writing tests can help provide such information. Writing tests can indicate the probability that a given student will find success and satisfaction in a program or professional position for which writing skill is a prerequisite. More specifically, normative test data can help students, their parents and their guidance counselors answer students' typical questions: Should I pursue advanced training in a postsecondary educational program in which writing is a key element? In which school or job am I most likely to be successful? Though test scores should never serve as the sole basis for answering such questions, they can play a valuable role.

Using Tests to Select Students

Admission. It is not uncommon to have more candidates than program openings. When this happens, teachers, counselors and administrators must select students for admission. Performance indicators such as writing tests can be used to rank order examinees to facilitate selection. Selection decisions most often affect those at either end of the skill continuum. That is, more able students are selected for inclusion in advanced writing programs, while less able students are selected for remedial writing programs.

Certification. Tests tailored to a specified certification domain are often used to verify and document a student's mastery of specific knowledge or skills. For example, teachers might use writing tests to certify mastery of beginning writing skills for purposes of grading or promotion. Or district and state administrators might use minimum writing competency tests as criteria for high school graduation. Both examples show how certification may be accomplished through testing.

Using Tests to Evaluate Programs

Survey Assessment. Survey assessment refers to the collection of group achievement data to determine general educational development (e.g., in writing). Data may be gathered by administering a writing test to a carefully selected random sample of students in the target population. Survey assessment is often cyclical, thus allowing for the examination of trends in writing skill development over time. Decision makers include (1) building-, district- or state-level administrators who allocate resources for special instructional needs pinpointed by the assessment, or (2) the public, which makes value judgments regarding perceived and reported levels of student writing skill development.

Formative Evaluation. In the context of formative program evaluation, program administrators and teachers attempt to determine which components of instruction are functioning as intended and which need further refinement. They may test students on each of the intermediate and final outcomes of a writing program, for example. Assessment for formative evaluation may also involve multiple test administrations to determine the effectiveness of ongoing modifications in a writing program.

Summative Evaluation. Summative evaluation reveals a program's overall merit, suggesting whether that program should be continued or terminated. Tests designed to assess students' performance on final learning outcomes are an important part of such an evaluation. Teachers, program, building or district administrators, and the public (including the board of education) may be involved in summative evaluation decisions. As with survey assessment and formative evaluation, multiple test administrations are common. Tests may be given prior to as well as following instruction, with retention testing after a given time interval.

Selecting Examinees as a Function of Purpose

In the three program evaluation contexts just cited (survey assessment, formative evaluation, and summative evaluation), testing costs can be significantly reduced through random sampling. If the student population is very large, then data summarized across a carefully selected random subset of students will reflect group performance every bit as accurately as if every student were tested—often at a fraction of the cost. It is not within the scope of this paper to present all the important considerations in sampling, as each specific educational situation is unique. The intent is to point out the potential financial advantage of sampling and to urge its consideration

It should be apparent that sampling is not feasible with instructional management or student screening decisions because in these contexts, individual student data are necessary.

Developing Exercises as a Function of Purpose

Generally, the process for developing writing assessment exercises remains constant across all eight educational assessment contexts. Careful planning is essential in all cases, and attention must always be given to designing exercises that give the examinee sufficient opportunity (in terms of time, appropriate stimulus and range of tasks) to demonstrate proficiency. Further, in all cases, the type of audience and purpose for communication should be made clear to the student. In addition, exercises should frame challenging tasks based on varied and directly relevant stimulus materials. And finally, in all cases, clear and concise instructions are essential.

A few factors vary according to context and the nature of the decisions to be made. As a general rule, the specificity of an exercise (i.e., level of detail in instructions) should increase along with the specificity of the skills to be assessed. In other words, exercises to be used in broad survey assessment need not be quite so focused as exercises to be used in, say, a diagnostic test.

The amount of writing required might also vary, depending on the decisions to be made. For example, it might be possible to rank order students in terms of general writing proficiency (via holistic scoring) on the basis of three or four general, relatively short writing samples. However, it would probably he very difficult to use those same three or four short writing samples to rehably and validly determine whether a student had mastered 10 to 15 specific, independent writing skills. Generally the more precise and numerous the criteria and standards of acceptable performance, the more writing neeaed to evaluate performance.

And finally, exercises developed for use in a large-scale statewide assessment or where important selection decisions are pending *must* be (1) independently reviewed by writing and assessment experts and (2) pretested. Pretesting and review are less critical with writing assessment exercises used in inst uctional classroom management.

Selecting Sccring Procedures as a Function of Purpose

Selection of scoring procedures is, in effect, part of assessment planning, since this decision is influenced by the purpose for the assessment and criteria to be used in judging writing proficiency. Though it is possible to conceptualize instances within each of the eight educational assessment contexts in which any given scoring approach could be employed, the actual scoring approach most commonly used will vary by context.

To illustrate, **diagnosis** of individual student strengths and weaknesses demands the level of specificity provided through analytical, primary trait or mechanics scoring. **Placement and guidance**, on the other hand, may only require holistic ratings because the objective of assessment is simply to rank order students on a continuum of writing skill.

Consider measurement of student status. While **selection** may require a holistic ranking of students, **certification** may be done through holistic ratings or analytical or primary trait scoring, depending on the specificity of the minimum competencies to be certified.

Holistic scoring procedures are well suited to the relatively broad, unfocused nature of large-scale survey assessment. However, analytical scoring may serve as well if the desire for

Table 4

Writing Assessment Procedures as a Function of Assessment Context

	Assessment Context		Assessment Procedure	
Context	Decision to be made	Decision makers	Examinees assessed	Exercise specificity
Diagnosis	Determine and track educational development	Teacher Student	Individual	Specific
Placement	Match level of student development to level of instruction	Teacher Counselor	Individual	General
Guidance	Rank order for educational planning decisions	Administrator Counselor Teacher Parent Student	Individual	General
Selection	Rank order examinees for selection into instruction	Administrator Counselor Teacher	Individual	General
Certification	Determine mastery of specific competencies	Teacher Student	Individual	Specific
Survey Assessment	Policy decision re· status of student educational development	Administrators Public	Sample	General
Formative Evaluation	Determine components of instructional program in need of revision	Program Developer Teacher	Sample	Depends on program objectives
Summative Evaluation	Program continuation	Administrator	Sample	Depends on program objectives

	Assessment Procedure				
Context	Holistic	Analytical	Primary trait	Mechanics	T-unit
Diagnosis			X	X	X
Placement	X	X			
Guidance	X	X			
Selection	X	X			
Certification		X	X	X	X
Survey Assessment	X	X			
Formative Evaluation		X	X	X	
Summative Evaluation	X	X			

individual data justifies the additional time required.

Scoring procedures for **formative evaluation** depend on the specificity of the enabling and terminal objectives that guide instruction. If overall writing proficiency is the focus of the program, analytical scoring may be selected. However, if instruction focuses on situation-specific rhetorical skills, primary trait scoring may be most appropriate. Similarly, emphasis on mechanics indicates selection of a corresponding scoring approach. In most instances, formative evaluation demands scoring procedures more specific than holistic.

With **summative evaluation**, holistic assessment may provide sufficient data to judge program viability. However, if stated program goals subdivide writing skill into component parts, analytical scoring may be appropriate. Instructional programs in writing seldom focus on a single rhetorical circumstance Rather, they deal with writing of many types, for many purposes. Therefore, primary trait scoring will have limited value in this context.

Ensuring Efficient, Effective, and High Quality Assessment

The keys to successful direct writing assessment are careful planning, thoughtful and creative exercise development, and consistent application of performance criteria during scoring. If these factors are given meticulous attention, the assessment will yield data that are (1) sufficiently precise to support necessary decisions, (2) reliable, (3) valid for the intended purpose, and (4) maximally cost-effective.

The preceding discussion is intended to acquaint the interested educator with available assessment strategies and to highlight some of the issues involved in selecting a scoring procedure appropriate for a specific context. Table 4 provides an overall summary of the key points made in that discussion

The reader is encouraged to refer to the list of REFERENCES *following this section and to the* APPENDIX, *which names contact persons in many states who can offer further information on writing assessment approaches and contingencies. In addition,* CAPT *welcomes further inquiries regarding writing assessment*

REFERENCES

American College Testing Program. Alternative strategies for the assessment of writing proficiency. Iowa City, IA: Author, 1979.

Braddock, R., Lloyd-Jones, R., and Schoer, L. *Research in written composition*. Urbana, IL: National Council of Teachers of English, 1963.

Breland, H.M., Conlon, G.C., and Rogosa, D. *A preliminary study of the writing ability*. New York, NY: College Entrance Examination Board, 1966.

Cronbach, L.J. Test validity. In R.L. Thorndike, *Educational Measurement*. Washington, D.C.: American Council on Education, 1971.

Diederich, P.B. Measuring growth in English. Urbana, IL: National Council of Teachers of English, 1974.

Fredrick, V. Writing assessment research report: A national survey. Monograph published by the Wisconsin Department of Public Instruction, Madison, WI: 1979.

Godshalk, F.I., Swineford, F., and Coffman, W.E. *The measurement of writing ability*. New York, NY: College Entrance Examination Board, 1966.

Hogan, T.P., and Mishler, C. Relationships between essay tests and objective tests of language skills for elementary school students. *Journal of Educational Measurement*, 1980, *17*, 219-227.

Hunt, K.W. Early blooming and late blooming syntactic structures. In C. Cooper and L. Odell (Eds.), *Evaluating writing*. Urbana, IL: National Council of Teachers of English, 1977.

Huntley, R.M., Schmeiser, C., and Stiggins, R. The assessment of rhetorical proficiency: The role of objective tests and writing samples. Paper presented at the annual meeting of the National Council on Measurement in Education, 1979.

Melton, V., and McCready, M. Survey of large-scale writing assessment programs. Ruston, LA: Louisiana Technological University, 1981.

Moss, Pamela A., Cole, Nancy S., and Khampalikit, Choosak. A comparison of procedures to assess written language skills in grades, 10, 7 and 4. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA: 1981.

Mullis, I. The primary trait system for scoring writing tasks. Denver, CO: National Assessment of Educational Progress, 1974.

National Assessment of Educational Progress. *Writing mechanics 1969-1974: A capsule description of changes in writing mechanics*. Denver, CO: Author, 1975.

Page, E.G. *The analysis of essays by computer* (Final Report, U.S. Office of Education Project 6-1318). Storrs, CT: University of Connecticut, 1968.

Rivas, F. *Write/rewrite: An assessment of revision skills* (Writing Report No. 05-W-04). Denver, CO: National Assessment of Educational Progress, 1977.

Steele, J. The assessment of writing proficiency via qualitative ratings of writing samples. Paper presented at the Annual Meeting of the National Council on Measurement in Education, 1979.

Stiggins, R.J. A comparison of direct and indirect writing assessment methods. *Research in the Teaching of English* (in press)

APPENDIX: Large-Scale Writing Assessment Program Profiles

(from Melton & McCready, 1981)

TESTING METHOD

State	Grades Tested	Sample Size (X 1000)	Objective Test	Writing Sample	Exercises Developed By
Alabama	3,6,9	40-60*		X	State Department University Faculty Teachers
Delaware	1-8,11	<5	X	X	State Department NAEP Exercises
California	3,6,12	<5	X	X	Local Districts
Florida	3,5,8,11	<5	X	X	State Department Teachers
Hawaii	4,8,11	<5		X	Committee
Idaho	9	10-20*		X	State Department
Louisiana	3,7,10	<5	X	X	Teachers
Maine	4,8,11	10-20*		X	NAEP Exercises
Maryland	9-12	>60*	X	X	State Department Contractor Teachers Local Districts
Massachusetts	7,8,9,12	>60*		X	State Department Teachers

*Entire Population Tested

WRITING SAMPLE DESCRIPTION

Kind of Writing	Scoring Method	Results Used By	Contact
Narration	Holistic	Local Districts Schools Teachers	Dr William Berryman State Dept of Education Room 607, State Office Bldg. Montgomery, AL 36130
Narration Persuasion	Primary Trait	Local Districts Schools Teachers State Department Public Report	Mr. Robert Bigelow State Dept. of Public Instr Townsend Bldg., Box 1402 Dover, DE 19901
Varies by District	Holistic	Local Districts Schools Teachers	Dr Dale Carlson State Dept. of Education 721 Capitol Mall Sacramento, CA 95814
Special Task	Analytical	State Summary Disseminated on request	Dr Thomas H. Fisher State Dept. of Education Knott Building Tallahassee, FL 32301
Narration Exposition Description Persuasion	Holistic	Local Districts	Dr Selvin Chin-Chance State Dept of Education Queen Liliuokalani Bldg 1390 Miller Street P O. Box 2360 Honolulu, HI 96804
Narration Exposition Description Persuasion	Holistic	Local Districts Schools	Ms. Karen Underwood State Dept of Education Len B Jordan Office Bldg Boise, ID 83720
Narration Exposition Persuasion	Primary Trait	Local Districts Schools Teachers State Department	Mr Joseph Williams Bureau of Assessment State Dept. of Education P O Box 44064 Baton Rouge, LA 70804
Narration Exposition Description	Holistic	Local Districts NAEP	Dr Horace P Maxcy, Jr State Dept. of Educational and Cultural Services State Office Building Augusta, ME 04333
Narration Exposition Description	Holistic	State Department Local Districts Schools	Dr William Grant State Dept. of Education BWI Airport P O Box 8717 Baltimore, MD 21240
Description Persuasion	Holistic Analytical	Local Districts Schools	Dr Allan Hartman State Dept of Education 31 St James Boston, MA 02116

TESTING METHOD

State	Grades Tested	Sample Size (X 1000)	Objective Test	Writing Sample	Exercises Developed By
Michigan**	4,7,10	<5		X	State Department University Faculty NAEP Exercises
Minnesota	1	<5		X	State Department Teachers University Faculty Local Districts
Nevada	3,6,9-12	5-10*	X	X	State Department Teachers
New Hampshire	5,9,12	<5		X	State Department
New Jersey	9	>60*	X	X	Contractor
New Mexico	10	Unspecified		X	State Department Teachers Local Districts
North Carolina	11	<5		X	Contractor
Ohio	8,12	<5	X	X	State Department University Faculty Teachers
Oregon	4,7,11	<5	X	X	State Department
Pennsylvania	5,8,11	Unspecified	X		

*Entire Population Tested
**Assessment Under Development

WRITING SAMPLE DESCRIPTION

Kind of Writing	Scoring Method	Results Used By	Contact
Narration Exposition Description	Primary Trait	To be specified	Dr Edward Roeber Michigan Dept of Education 620 Michigan National Tower P O Box 30008 Lansing, MI 48909
Narration Exposition Description Persuasion	Primary Trait	Statewide Reporting	Dr William McMillian State Dept of Education Capitol Square, 550 Cedar St St Paul, MN 55101
Exposition Description Persuasion	Holistic	Local Districts Schools Teachers Parents & Students	Dr R Harold Mathers State Dept of Education 400 West King Street Carson City, NV 89701
Narration Exposition	Holistic	State Department Local Districts	Dr. James V Carr State Dept. of Education 64 North Main Street Concord, NH 03301
Narration	Holistic	Local Districts School Teachers Students	Dr Stephen Koffler Department of Education 225 West State Street Room 200 Trenton, NJ 08625
Description Persuasion	Holistic	Local Districts Schools Teachers	Dr Carroll L Hall State Dept of Education Education Building Santa Fe, NM 87503
Unspecified	Holistic Analytical	To be specified	Dr William J Brown State Dept of Public Instruction Raleigh, NC 27611
Narration Exposition Description Persuasion	Holistic	State Department Local Districts Schools	Mr Jim Payton State Dept of Education 65 South Front Street Room 804 Columbus, OH 43215
Narration Exposition Description Persuasion	Holistic	State Department	R B. Clemmer Oregon Dept of Education 700 Pringle Parkway SE Salem, OR 97310
		State Department	Dr Robert Coldiron State Dept of Education P O Box 911 Harrisburg, PA 17126

TESTING METHOD

State	Grades Tested	Sample Size (X 1000)	Objective Test	Writing Sample	Exercises Developed By
Rhode Island	4,6,8,10	<5	X	X	Contractor
South Carolina	6,8,11	>60*		X	State Department University Faculty Contractor Teachers Local Districts
Texas	3,5,9	>60*	X	X	State Department University Faculty Contractor Teachers
Wyoming	6,9	<5		X	State Department University Faculty Teachers
City					
Little Rock, AR	1-11	Not specified	X		
Phoenix, AZ	9-12	5-10	X	X	Teacher Local Districts
Monterey, CA	1-12	<5		X	Teacher
Tallahassee, FL	1-8	5-10	X	X	University Faculty Teachers Local Districts
Atlanta, GA	1-12		X		
Des Moines, IA	9	<5		X	Teacher Local Districts

*Entire Population Tested

WRITING SAMPLE DESCRIPTION

Kind of Writing	Scoring Method	Results Used By	Contact
Narration Persuasion	Holistic	Local Districts Schools Teachers State Department	Ms Martha Highsmith State Dept of Education 199 Promenade Street Suite 204 Providence, RI 02908
Narration Exposition Description Persuasion	Holistic Analytical	Local Districts Schools Teachers State Department	Dr Vana Meredith State Dept of Education 1429 Senate Street, Room 604 Columbia, SC 29201
Narration Description Persuasion	Holistic	Local Districts	Mr Keith L Cruse Texas Education Agency 201 East 11th Street Austin, TX 78701
Narration Description	Holistic	Local Districts Schools	Dr Mark Fox State Dept of Education Hathaway Building Cheyenne, WY 82002
		Local Districts Schools Teachers	Dr Carolyn Weddle Little Rock School District West Markham & Izard Little Rock AR 72201
Narration Exposition Description	Analytical	Schools Teachers	Mr Gerald De Grow Phoenix U H S District 210 2526 W Osborn Rd Phoenix, AZ 85017
Exposition Description	Holistic	Schools Teachers	Dr Lloyd Swanson Monterey Peninsula Unified School District P O Box 131 Monterey, CA 93940
Narration Exposition Description	Analytical	Local Districts Schools Teachers	Mr F W Ashmore Leon Co Public Schools P O Box 246 Tallahassee, FL 32302
		Schools Teachers	Mr Alonzo Crim Int School District 203 224 Central Avenue S W Atlanta, GA 30303
Persuasion	Holistic Analytical	Schools Teacher	Mr Dwight M Davis Des Moines Int Comm Dist 1800 Grand Avenue Des Moines, IA 50307

TESTING METHOD

City	Grades Tested	Sample Size (X 1000)	Objective Test	Writing Sample	Exercises Developed By
Chicago, IL	9-12	>60		X	Teacher
Boston, MA	2,5,8	5-10		X	State Department Teachers
Wichita, KS	K-12	10-20		X	Teachers Coord of L A
Baltimore MD	1-9	>60	X	X	Teachers Local Districts
Detroit, MI	10-12	10-20	X	X	Contractor Dept /L A
Raleigh NC	1-12	<5	X	X	Contractor Teachers Local Districts
Albuquerque NM	1,6,9-12	5-10	X	X	State Department Teachers
Santa Fe NM	7-12	<5		X	Teachers
New York NY	8,11	<60		X	State Department
Portland, OR	3-9	Not specified	X		

WRITING SAMPLE DESCRIPTION

Kind of Writing	Scoring Method	Results Used By	Contact
Narration Exposition Description Persuasion	Analytical	Local Districts	Mr James Redmond Cook Co Public Schools 228 North La Salle Street Chicago, IL 60601
Narration Exposition Description	Holistic	Local Districts Schools Teachers	Mr William Leary Boston Public School Dist 15 Beacon Street Boston, MA 02108
Narration	Holistic	Local Districts Schools	Dr Alvin E Morris Wichita Sedgwick Unfd Dist 259 428 S Broadway Wichita Falls, KS 67202
Narration Persuasion	Analytical	Schools Teachers	Mr Roland Patterson Baltimore Co Public Schools 3 E 35th Street Baltimore, MD 21218
Exposition	Holistic		Mr Charles Wolfe Wayne Co Public Schools 5057 Woodward Detroit, MI 48202
Narration Exposition Description	Teacher Option	Schools Teachers Parents Students	Mr C L Hooper Raleigh Dist Public Schools 601 Devereux St Raleigh, NC 27605
Exposition Description Persuasion	Holistic	Local Districts Schools Teachers State Department Reported to Media Report to Student	Mr E Stapleton Bernalillo Co Public Schools Box 1927 Albuquerque, NM 87103
Description	Holistic Analytical	Schools	Mr Philip Bebo Santa Fe Co Public Schools 610 Alta Vista Santa Fe NM 87501
Exposition Persuasion	Holistic	Local Districts Schools Teachers State Department	Mr Calvin E Gross New York City Schools 110 Livingston Street Brooklyn, NY 11201
		Local Districts Schools Teachers Parents Students	Dr Walter Hathaway Portland Public Schools P O Box 3107 Portland, OR 97208

TESTING METHOD

City	Grades Tested	Sample Size (X 1000)	Objective Test	Writing Sample	Exercises Developed By
Austin. TX	3,9	<5	X	X	State Department University Faculty Contractor Teachers
Madison, WI	5,8,11	<5	X	X	State Department University Faculty Parent/Bus People
Seattle, WA	3,6,9-11	<5		X	Teachers Curr Specialists
Laramie, WY	6,9	<5		X	Com of local and state univ members

WRITING SAMPLE DESCRIPTION

Kind of Writing	Scoring Method	Results Used By	Contact
Narration Exposition Persuasion	Holistic	Local Districts Schools Teachers State Department	Dr Jack Davidson Austin. ESD 6100 N Guadalupe Austin. TX 78752
Narration Exposition Persuasion	Holistic Primary Trait	Not specified	Mr D S Ritchie Dane Co Public Schools 545 W Dayton Madison. WI 53703
	Analytical	Schools	Mr Forbes Bottomly Seattle School Dist 1 815 Fourth Ave N Seattle. WA 98109
Exposition Description	Holistic	Local Districts Schools State Department	Dr Joe Lutjeharms Laramie Co Public School District 1 Cheyenne. WY 82001