DOCUMENT RESUME

ED 183 619 TM 800 038

TITLE	Manual for Project Management. MEC-ETS Standard Setting Studies Conducted as a Part of MEC's Improving Basic Skills Project.
INSTITUTION	Educational Testing Service, Wellesley Hills, Mass.
SPONS AGENCY	Merrimack Education Center, Chelmsford, Mass.
PUB DATE	Apr 79
NOTE	104p.
AVAILABLE FROM	Merrimack Education Center, 101 Mill Road, Chelmsford, MA 01864 ($6.00)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Academic Standards; *Basic Skills; *Cutting Scores; Data Collection; Educational Objectives; Elementary Secondary Education; Essay Tests; Grade 3; Grade 6; Grade 9; *Guidelines; *Mastery Tests; *State Programs; Student Testing; Test Bias; Test Interpretation; Test Selection
IDENTIFIERS	Contrasting Groups Method; Holistic Approach; Massachusetts; Nedelsky Method

ABSTRACT

This document is designed to assist school districts to implement the Massachusetts State Board of Education regulations requiring the setting of performance standards in basic skills at three grade levels. The appropriate procedures are described in the manual: (1) setting of basic skills objectives, at three grade levels; (2) the selection of tests or other measures that adequately cover the objectives; (3) the setting of a minimum acceptable score on each test or measure; and (4) the reporting of the number and characteristics (sex, race, etc.) of those students achieving below the local standard. Information is appended on tests currently available for assessing basic skills development at the high school level and general achievement test series, as well as on holistic scoring. (Author/CTM)

MANUAL FOR PROJECT MANAGEMENT

MEC - ETS

STANDARD SETTING STUDIES

CONDUCTED AS A PART OF

MEC'S

IMPROVING BASIC SKILLS PROJECT

Second Draft

April, 1979

Produced for the

Merrimack Education Center

By

Educational Testing Service
New England Office

INTRODUCTORY NOTE

This second edition of the project managers' manual for the MEC-ETS standard setting study includes several changes based on the review of the earlier draft by Dr. Michael Zieky (ETS, Princeton). The earlier draft was also reviewed informally by Massachusetts Department of Education staff.

Since the publication of the first draft, the MEC-ETS project has proceeded on schedule, with a minimum of difficulty. The steps outlined in the manual have met this first test of feasibility.

George Elford
April, 1979

TABLE OF CONTENTS

Section 1

Overview of the Basic Skill Improvement Policy - Implications for School Districts

In January 1979, the Massachusetts State Board of Education published the regulations for implementation of the Basic Skills Improvement Policy adapted by the Board in August, 1978. This manual has been developed to assist school districts in carrying out the State Board's mandate which includes as one of the required steps, the setting of performance standards in basic skills at three grade levels.

While the manual will deal with the setting of performance standards, it is important to note that the setting of standards is only a means to an end. The purpose of standard setting and testing is to identify those students (the target group), who need special attention with basic skills. The purpose of the state's policy is to assure that these students are identified and that an effective program is provided for them. To this end, the state has called for standard-setting, testing, documented program planning, and a monitoring of program effectiveness - all with broad based community involvement.

At the outset, some clarification is needed on the meaning of the term "minimum standards" in the context of the state's regulations. The state regulations speak of "minimum standards" as a set of objectives in the basic skills and a level of achievement on the basic skill objectives. The objectives represent the content - what has to be attained by the minimally competent student. The level of achievement represents the degree to which the objectives must be attained in order for the student's skill level to be considered sufficient or minimally acceptable. The level of achievement can only be identified by some measured performance

level of achievement can only be identified by some measured performance which indicates a level of achievement. In effect, standards can become operational only when they are related to some measurement of the performance of individuals or groups. Any description of levels of achievement on certain basic skill objectives must be translated into a passing score on a test or a set of measures. Thus, for present purposes, the operational definition of a minimum standard is the minimum acceptable total score on an appropriate test or measure, which covers basic skills objectives in reading, writing, and mathematics.

To carefully implement the state regulation on standard setting, the following steps are called for:

1) The definition by the local staff of objectives, at three grade levels, that are considered basic skill objectives. At the secondary level, these objectives must include but can go beyond the list of objectives developed by the state.

2) The selection of tests or other measures that adequately cover the above objectives.

3) The setting of a minimum acceptable score on each test or measure. This score represents the minimum standard of basic skill development to which the local district chooses to hold itself accountable. Students achieving below this standard will be given special attention in the system's basic skills improvement program.

4) The reporting of the number and characteristics (sex, race, etc.) of those students achieving below the local standard.

While certain students will be identified as below the local standard based on their performance on a specific test or assessment exercise, local school officials need not conclude from this single test score that a given student is clearly in need or not in need of remediation. The more defensible use of test data calls for the tentative classification of a student as needing remediation based on a below standard and perhaps a borderline performance on the basic skills measure, with the final clarification depending on the corroboration of this evidence by other information. To cite an extreme example, a student who for the past three years scored in the 80th percentile in mathematics should not be classified as a below standard student when he or she on a given day did badly on one test. Presumably, these new regulations do not call for the abandonment of established principles in test interpretation. The cardinal principle in the interpretation of test scores is that test scores should be interpreted together with other available information - never in isolation.

Section 2

Local Decision-Making in Standard Setting

As noted earlier, the state regulations call for community involvement at several points in the planning and implementation of the local districts basic skills improvement program. Community involvement is required in the following:

1) the establishment and periodic review of program plans at all of the three levels described earlier
2) the establishment and periodic review of minimum standards for each basic skill at each level.

Community involvement includes participation of parents of public school students, including Title I parents, employers, students at the secondary level, teachers, administrators, and representatives of the general public.

In the implementation of this requirement, consideration should be given to efficient use of people's time as well as the expertise required for each of the tasks. The approach taken in this manual calls for leadership from teachers and school administrators in drafting proposed plans and formulating proposed standards with the community representatives joining with the staff on a systemwide advisory committee, which will make recommendations to the school committee on both the basic skills program and the local performance standards.

The following table depicts the roles of these two groups at each point in the process:

		Professional Staff	Advisory Committee
1)	The identification of basic skill objectives	initial recommendation	final recommendation
2)	Selection of appropriate tests	completed by staff	reported to committee
3)	Setting of performance standards on tests or other measures	initial recommendation based on staff judgement and field testing.	final recommendation
4)	Planning instructional programs for target population	initial draft of plan	final recommendation

The advisory committee,* as shown above, does not make final decisions but only final recommendations. Final decisions are the designated responsibility of the local school committee. In the approach presented above, the professional staff assumes the responsibility for the actual selection of tests, which entails the matching of test items to the basic skill objectives. This is a technical activity requiring some familiarity with tests and objectives. In the setting of standards, the procedure recommended here begins with staff judgement, because the educational standards maintained by the school, for example in grading, promotion, and placement decisions are found in the consensus of teacher judgement.

The advisory committee referred to above, as envisioned here, would be a system-wide committee made up of staff and community representatives as called for in the state regulations. The work of professional staff described above would take place in three grade level committees. Needless to say, already existing committees and local conditions will undoubtedly shape the final design of the decision-making structure in each community.

*The advisory committee represents simply one approach to community involvement; surveys, public hearings, etc., represent other possible approaches.

The Merrimack Regional Planning Council will direct the design and implementation of the services made available, for the purposes of implementing these regulations, to those districts served by the Merrimack Education Center. This council will be made up of one representative from each district participating in the project.

Section 3

Identification of Grade Levels and Basic Skill Objectives

The school committee is called upon to designate three grade levels at which basic skills testing and standard setting are required. The project facilitators will undoubtedly meet with the appropriate administrative and supervisory staff to identify the specific grade levels to be proposed to the school committee. The regulations describe the three levels as follows:

1) Early elementary - K through grade 3

2) Late elementary - grade 4 through grade 6

3) Secondary - grade 7 through grade 12 (with testing begun no later than grade 9)

At each level, the local school district must a) identify students in need of special attention in basic skill development by administering tests for which a local performance standard has been set, and b) develop, maintain, and periodically evaluate a program to meet the needs of these students.

After the three grade levels have been determined, the staff needs to identify the skill content in reading, writing, and mathematics that they judge to be basic for students at each of these grade levels. A staff committee at each grade level should examine the present curricular objectives (and, if need be, other objectives) and select those which they consider basic. At the secondary level, the objectives identified by the local district must include but may go beyond those objectives specified in the state regulations.

The selection of objectives at the two elementary levels might well begin with the identification of those objectives which at the elementary

level represent reasonable steps toward the attainment of the state specified secondary level objectives. Many of these secondary level objectives can be attained to a limited degree in elementary grades. The successful attainment of complex secondary level objectives clearly depends on the earlier attainment of more elementary objectives. As a guiding suggestion, grade level committees might do well to begin by reviewing the state specified secondary level objectives because the attainment of these objectives by virtually all of the students is the ultimate goal of the basic skill improvement program. Perhaps a chart showing where each of these final objectives is introduced and attained by most students might prove useful. This chart might, for example, show the following in reference to one basic skill objective in mathematics:

Early Elementary	Later Elementary	Secondary (State Req.)
1.) Add and subtract whole numbers (one and two-digit numbers)	1) Add, subtract, multiply and divide whole numbers (one and two digit numbers)	1) Add, subtract, multiply and divide whole numbers (three to five digit numbers).

The end product of this first task by each of the grade level committees is a complete list of grade level basic skills objectives in reading, writing, and mathematics. This list should not include all objectives given attention at this level but only those considered basic. Presumably, the more able students will be working on learning activities related to more advanced objectives at each of these grade levels. It is important that these minimum objectives do not become the maximum. The state requirement deals with the attainment of minimum competency in basic skills by virtually all of the students. It does not describe the full range of objectives to be attained

by the average and above average students. To make clear this distinction between the minimum objectives in basic skills and the full range of expected attainment, the grade level committees might do well to specify those objectives which they do consider not as basic skills as covered by the state program but as more advanced skills which they nevertheless expect a large number, if not the majority, of students to attain.

The lists of basic skill objectives identified by each of the three grade level committees would then be presented for review and approval by the systemwide advisory committee. These lists of objectives will then guide the selection of the tests to be administered at the three grade levels. These tests should provide substantial coverage of most if not all of these objectives. These lists will also guide the description of the learning activities which will, in fact, constitute the basic skills improvement program.

Section 4

Test Review and Selection

In matching tests to listed basic skill objectives, two possible approaches are either the selection of tests already available or the development of new tests to meet local objectives. Given the present state of the educational economy, the latter course, local district test development, should be embraced only as a last resort. For all but the largest school districts, the unit cost of such a locally developed test with a scoring and reporting system, would in most cases be prohibitive.[1] This manual will assume that at all three levels, MEC districts will be using currently available tests.

Appendix B includes a comprehensive but by no means exhaustive list of basic skill tests available (as of January, 1979). As a first, rough cut in selecting tests, grade level committees should identify tests that meet the following general requirements:

1) Test content is appropriate, i.e., covers basic skills based on general descriptions provided by the publisher

2) The kinds of items used are appropriate based on general descriptions and examples provided by the publisher

3) Test has adequate statistical properties, e.g., reliability of total score at .85 or above.

4) Useful score reports are available, which include:

   a) a frequency or local percentile distribution which is necessary to determine the impact of a proposed standard

   b) norm referenced percentile scores, which are extremely useful (but not absolutely ne-essary) in reviewing a proposed standard

[1] A future service offered through MEC might include an item bank of basic skill items, pre-tested and scaled by different levels, and a uniform scoring and reporting service.

c) cluster scores and/or individual student profiles, which are useful for subsequent instructional planning

Tests which fail to meet the above requirements should be set aside.[1] Tests that do meet these requirements should then be subjected to additional scrutiny by the appropriate grade level committees. These committees will undoubtedly choose to review first of all those tests currently in use which meet the above requirements. All tests should then be reviewed item by item by one or more (preferably 2 or 3) committee members using the Basic Skills Item Review Form (see Appendix C). In completing the review form, the committee members rate <u>each item</u> according to:

1.) when the content covered by the item is taught in this school system in reference to the grade level in question - earlier? later?

2.) whether the content covered by the item is basic or advanced for this grade level?

3.) which basic skill objective does the item measure? (Cite the objective by number from the local list).

As a rule of thumb, any test in which three fourths or more of the items are judged as measuring basic skills appropriate to the grade level would be acceptable for present purposes, provided the test also covers these objectives in a thorough and balanced fashion. A test presently in use that satisfactorily meets the above requirements could be chosen for the sake of economy and other useful purposes over a test that had <u>slightly</u> better coverage of the basic skill objectives.

[1] In selecting tests at the secondary level, it will also be necessary to check with the state's list of approved tests, which will be available in May, 1979.

Given the nature of these tests and basic character of the basic skills as described here; there is little danger of students excelling on advanced items while doing poorly on the basic skills items on the same test. An analysis of error patterns from time to time would answer any questions concerning this possible problem. For example, on a 50-item test in which 35 items were identified as basic skill items and 30 correct answers was the minimum acceptable score (the standard), one would assume and could readily verify that those performing below the standard missed most of the advanced and some of the basic skill items. One would also assume and could verify that those above the standard missed very few of the basic skill items and some of the advanced items. Admittedly, this argument rests upon some assumptions that can and should be verified.

Having identified basic skill objectives and selected tests to match these objectives, the grade level committees are prepared to begin setting a standard or cutting score based on local expectations.

The Scoring of Writing Samples

The discussion of testing thus far in this manual has dealt with objective, machine scorable tests. The state requirements call for the scoring of a writing sample at the secondary level (and perhaps at one or both of the elementary levels). The scoring of writing samples requires a set of procedures which enable each writing sample to be scored in a consistent manner. This could be accomplished by a set of detailed decision rules which left no room for teacher judgement. For example, one could devise a detailed analytical scoring procedure in which one point is taken off for every word misspelled, one point off whenever subject and verb do not agree, etc.

A second method, called holistic scoring, provides for the consistent scoring of writing samples by having these writing samples scored in a workshop setting in which teachers come to consensus understanding of a scoring system in which a single score is assigned on the basis of a complex set of criteria. These criteria are formulated only in a very general way. In the holistic approach, consistency is assured because each piece is judged by more than one reader, with inconsistencies in those judgements resolved by additional readings. (Appendix D describes some of the features of holistic scoring.)

As a part of the MEC project, an essay scoring workshop will be offered on May 17 for all participating districts. At this workshop only secondary level papers will be scored. Additional sessions might be added for other levels.

The tasks required for a holistic essay scoring are first listed here and then described in more detail below.

1) the selection of the topic or writing exercises

2) the administration of the writing exercise

3) the selection of readers for the scoring session

4) the submission of completed writing samples to MEC

5) the additional coding of papers by MEC (system, school, teacher codes)

6) the selection and duplication of training papers

7) the conduct of the scoring session

8) the setting of a proposed standard for each district

9) papers returned to schools (teachers)

10) scores decoded and class roster lists prepared

One end product of this process will be a list of students, each with a writing score. These scores, along with other test score data, will be collected for computer analysis as part of the standard setting study.

Several of the tasks listed above call for further explanation.

1) The selection of the writing exercise. With assistance from the ETS Center for the Assessment of Writing, MEC will make available a list of possible writing exercises, from which each district's grade level committee will rank the exercises in order of their preference. The exercise which enjoys the highest total ranking from all districts will be the one used by all participating districts. It is essential that all districts involved in the same scoring session use the same writing exercises.

2) The administration of the writing exercise. The student essays are to be written in the classroom without aids such as a dictionary. Allow twenty minutes and twenty minutes only. The date of the assignment is not restricted; however, MEC must receive papers by Tuesday, May 8. The students may write on any type of 8½ by 11" paper. Both sides of the paper may be used, though it

is important that each student use no more than one piece of paper.

Before the essay writing, each student receives 3" x 5" card or other paper in addition to the paper on which the essay will be written.

Each student records his or her name, teacher, class, and school on the card. At the bottom of the card, the student writes a five digit number followed by a single letter that he or she has created. (The only restriction is that each digit of the number must be different.) For example, 92831G, or 47682Z. This code is used to assure anonymity of essays during the essay reading workshop.

The student then copies the number and letter exactly as above at the upper right corner of the essay paper. Students should be told not to put any other identifying information on the essay paper.

The teacher is to collect the cards and retain them. This will enable the teacher to decode the papers upon return from MEC, so that papers can be returned to the students and score roster lists prepared.

After the essay writing, the essays are collected by the teacher. Using the postage-prepaid labels the teacher forwards to MEC the student essays with a cover sheet indicating school, teacher, grade, and class. The receipt of the essays at MEC must be guaranteed no later than Tuesday, May 8. A sample topic and teacher instruction sheet is included in Appendix E.

3) The selection of readers. The readers should be teachers of writing at the same grade level as the students whose papers are being scored (or supervisory staff familiar with student writing at several levels). For the one day workshop, one teacher (scorer) should be available for every 200 student papers submitted, if the student responses average about three quarters of a page in length. For every 10 readers at the scoring session, two aides (parents, older students, etc.) will be required.

4, 5, 6, & 7) The submission of the papers to MEC, additional coding, the selection of the training papers, and the conduct of scoring session will be provided for MEC by ETS. (For information on these steps, consult the ETS Basic Skills Assessment Manual for Scoring the Writing Sample.)

8) At the completion of the scoring session, the teachers who serve as scorers from each district will be asked to select a particular score that represents, in their judgement, the minimum acceptable performance on the writing sample in their district. They will then review this proposed standard in a subsequent meeting with the grade level committee, in their district, using sample (training) papers to illustrate what the scores mean. This proposed standard will be included in the final report to the Advisory Committee as one of the proposed standards.

9) MEC will re-group the essays by school and by teachers within each school for their return to the classroom.

10) Using the master code list provided by MEC (see Appendix F), each teacher will record each student's score on a roster list. A copy of this list will then be returned to the facilitator for inclusion in the data collection phase of the standard setting study.

These same procedures will be followed in the event that additional scoring sessions are scheduled for one or both of the elementary levels.

Section 6

Methods of Setting Performance Standards

Introduction

The main goal of this section is to describe two rational methods local school district personnel can use to arrive at standards on proficiency tests. There is no single method or procedure for setting standards that is good for all situations. When a standard is necessary, care should be taken in selecting an appropriate procedure. Further caution must be exercised in interpreting and using standards, particularly when the consequences of a student being above or below the standard are serious.

The Use of Minimum Performance Standards

A standard on a proficiency test can be used to define the level of mastery of a basic skill that must be reached by a student to satisfy locally determined standards. A standard is our best estimate of how much mastery is enough to be reasonably assured that a student has mastered those skills. A strength of using a standard lies in its impartial application to all students. However, when standards are capriciously arrived at or followed blindly, the desired virtues of impartiality can become indiscriminate or systematically unfair obstacles to students.

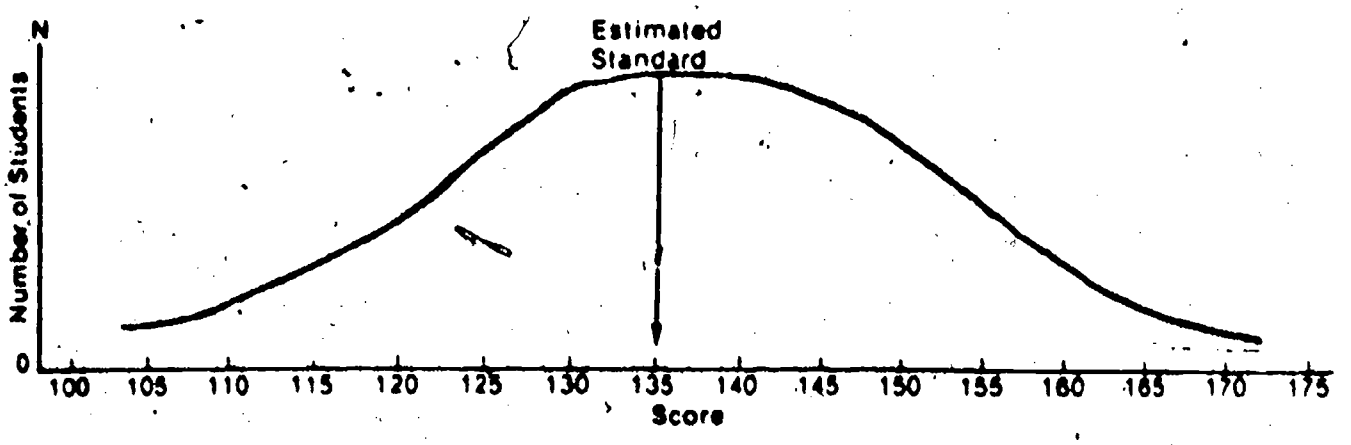
One dangerous implication of a standard is that it is often perceived as an absolute and precise indicator of who has achieved mastery. But tests are not perfect measurement instruments, students are not perfect test takers and standards are not set by perfect judges. Given

these important sources of error, there is very little assurance that a student with a score just above the standard is superior to a student just below the same standard. A standard, properly set, only can assure us that on the average we will make fewer mistakes using that score than using any other score.

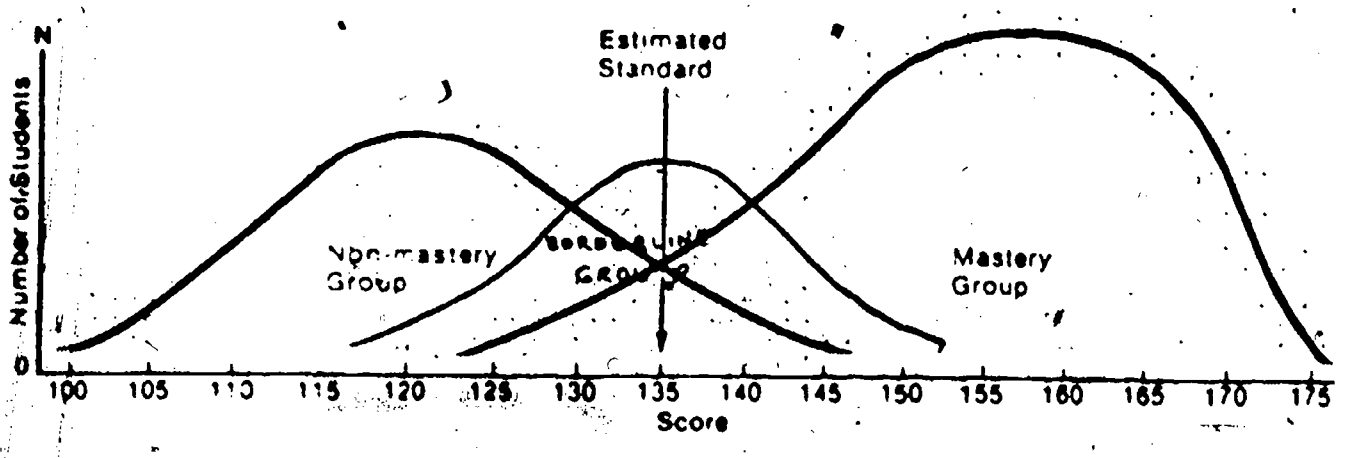
Therefore, in situations where test performance is an integral part of a district's requirements for student placement, standards should be determined with care and caution. In any case, a test score is but one piece of information about a student and must be interpreted in light of other data, including courses taken, classroom performance, teacher judgements, and other test data.

Types of Errors

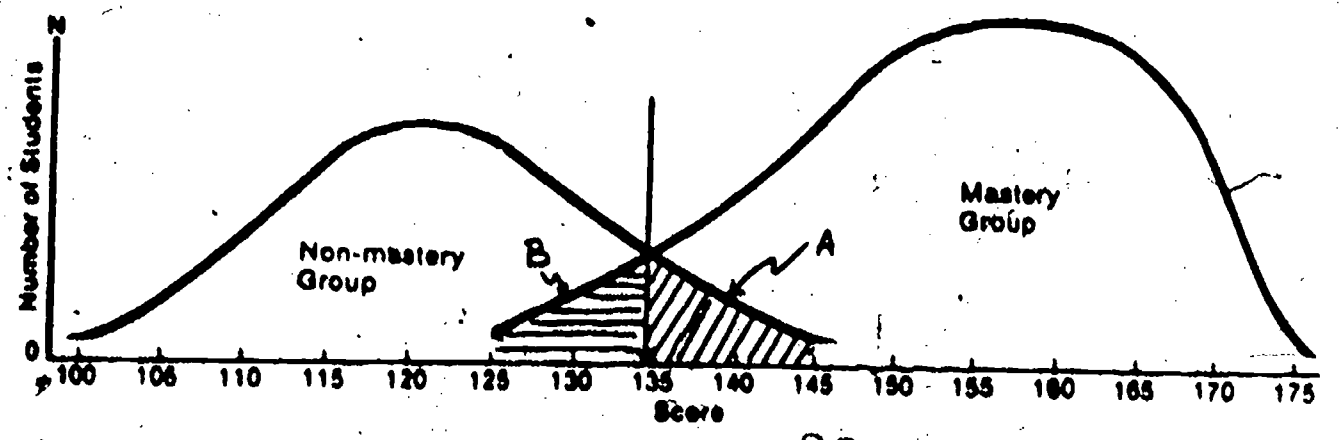
Ideally we would like to be able to identify students needing supplemental instruction without making any errors in classification. Our experience in administering tests to representative groups of students tells us that students' test scores tend to be normally distributed like below.



Most of the students will be "masters" of the material, some will be borderline students, and some number of students will be "non-masters". But given the imprecision of measurement (from both test itself and student performance factors) there is no place to draw a line (standard) through the distribution that separates the mastery groups without error. We could keep our classification errors to a minimum if the three groups of scores were distributed like those below, with the mastery and non-mastery groups clearly separated and the borderline students in the middle.

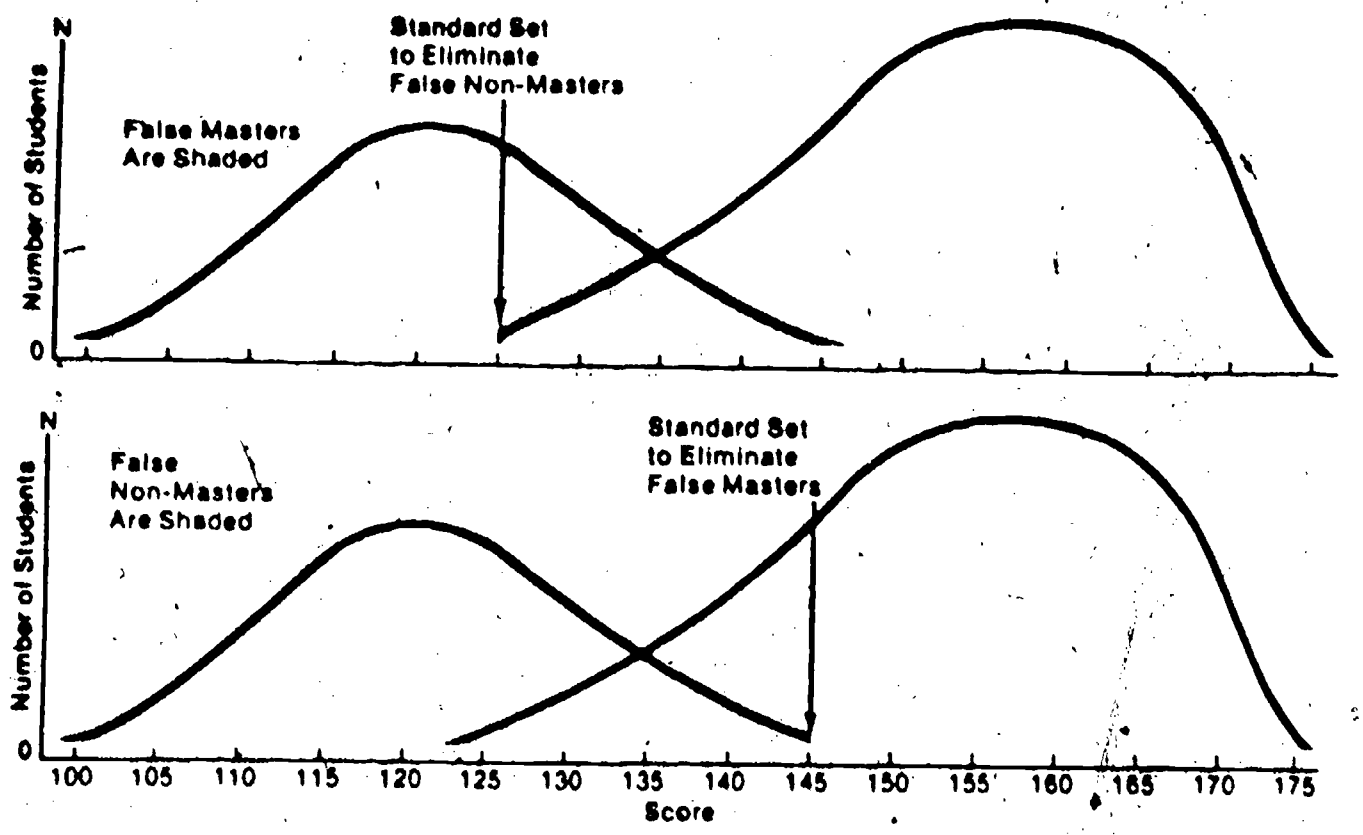


The performance-based methods of selecting standards allow us to look at student performance in just this way. The identification of the groups, however, adds another source of potential errors - teacher's judgements. More will be said later about how these procedures work. In practice the mastery and non-mastery groups tend to overlap as shown.



If the standard is at the vertical line in the figure we can see the approximate magnitude of two types of errors in classification. The students whose scores fell in the shaded part of the curve (A) are the non-masters" who "passed" the criterion. This group is sometimes referred to as "false-positives". The students whose scores are in the other shaded section (B) are the "masters" who "failed" to meet the criterion.

By moving the standard higher or lower along the horizontal axis, we can reduce one or the other kind of error. But when we reduce the errors of one type we increase those of the other type. Sometimes tradeoffs can be rationalized. As one moves the standard higher he is saying that it is more important to identify all or most of the "non-mastery" for supplemental instruction with the risk of including more students who have already mastered the skill in the supplemental program. As one moves the standard lower, one is willing to let more "non-masters" pass to avoid holding back any potential masters.



Although some errors in classification are inevitable, they should be held to a minimum. Consideration should be given to the direction of the errors the district would rather make. The following social, political and economic issues are legitimate concerns that affect the placement of standards.

- District resources for providing supplemental instruction. A district must have appropriate staff (both in number and training), space and program material to aid all students identified as needing supplemental assistance.

- Proficiency assessment as mandated by the state is not intended to be punitive to students. The intention is clearly on identification of students who need help in gaining mastery of the basic skills.

- Standards should be locally determined. Therefore they should reflect and be sensitive to the community's concerns. Input should be sought from representatives of all affected and interested groups in the community.

- The process of setting standards is new for most districts, so allow for mistakes. Standards should be reviewed every year. If standards are going to be made more stringent be sure to give sufficient notice to affected students and parents.

Types of Judgements

There are four ways of obtaining judgements to arrive at a standard.

- Judgement "uncontaminated by data"

- Judgement based on inspection of the instrument

- Judgement based on examination of student performance

- Judgement based on both inspection of the instrument and examination of performance data.

Unfortunately the first procedure is the one most frequently used and most difficult to defend. Examples include the historical reliance

on 70% correct as "passing" without reference to the test or the population. Similarly, standards set arbitrarily high without reference to either the test or student in order to appease a board member or "look tough" to competing schools are on shaky ground.

The second kind of judgement (item-review) takes into account several test variables including test content, appropriateness of objectives, and item difficulty. Direct inspection of the test, item-by-item, generally is made by a group of "experts", people who are familiar with the institutional and test objectives and knowledgeable about the test content.

The third set of judgements (performance-based) depends on how students perform on a test. In the simplest method, students are ranked and selected either from the top down or bottom up depending on how many can be accommodated. Performance-based methods can be used in a predictive sense as well as a normative sense. When judgements about the students' mastery of a subject are collected from knowledgeable people, standards can be set to reduce the liklihood of making classification errors.

Both item-review and performance are important factors if one is to make reasonable and fair judgements. The weakness of either approach alone is greatest when the judges lose touch with the interrelationships of the factors. In using the methods described here, it is important that judges not make decisions in a vacuum but keep in mind all the linkages between tests, student performance, instruction and district objectives.

Finally, standards obtained by any of these methods are not absolute decisions. They should be treated as recommendations. A final adjustment may be necessary by a review group or the school committee.

General Preparation for a Standard Setting Session Using Item-Review Methods

Identify and Select Judges:

The grade level committee might itself serve as the panel of judges. Prior to the work session at which the standard setting judgements are to be made, the facilitator should make the following preparations:

- Obtain enough copies of each test to be able to distribute one to each judge. (An overhead projector can be used, but it is better if judges have their own copy too.)

- Prepare enough copies of the Judge's Recording Forms for each judge (if more than one test is being inspected, confusion of paper shuffling can be reduced by duplicating the forms on different colored paper for each different test.)

- Make enough practice sheets for all of the judges.

- Have copies of the district's definition of minimal competence made for each judge.

- Arrange for a comfortable meeting room with sufficient table or desk space for the judges.

- Arrange for someone to assist you at the meeting.

- Obtain one copy of the technical manual and/or printout of data on item difficulty and typical student performance. Also have a copy of the answer key.

The Nedelsky Method (Method B)

The Nedelsky Method requires a group of judges to review a test and make a decision for each question on how many wrong alternatives a minimally qualified student would be expected to eliminate as being obviously wrong. The number of alternatives remaining constitutes the set from which that student is expected to guess. In this way each judge assesses the difficulty of every item. The overall difficulty (or standard) of the test is the averaged judged difficulty of all items. The procedure can be used only with multiple-choice tests.

The critical elements in determining a standard with the Nedelsky Method rest primarily in the selection of representative judges and reaching an understanding and agreement on what is meant by the "minimally competent" student.

It is recommended that between 7 and 9 judges be used with the Nedelsky Method. Factors to consider in selecting committees and judges are discussed elsewhere.

In defining the minimally competent, or just barely passing, student it may help to describe characteristics of such students from one's own experience and create a composite hypothetical profile as a model for discussion with other judges. Sometimes it is helpful to state what such a student is and is not like. For instance,

- The minimally competent high school graduate generally is not a college bound student.
- The difference between the "average" high school graduate and the minimally competent graduate may be as great as the difference between that average graduate and the brightest student getting a scholarship to a selective university.
- The minimally competent high school graduate is like (a locally known person or occupation) who works as a (some appropriate occupation) at (a local shop or company).

One source of confusion in defining this mythical student's abilities sometimes arises during the course of the Nedelsky procedure. The difference betwen what this student "should" be able to do and what this student "could" do can be difficult to resolve. For some judges, there is a tendency to expect all students "should" be able to get all the questions correct. For other judges, even many competent students "couldn't correctly answer many of the questions. The standards need to be realistic and reflect the content of instruction as it exists for the students. Yet, if there is dissatisfaction with the present standards, more might be expected of the students. These issues should be raised during the standard setting session. Even though everyone may not be in complete agreement it is important that the various views be presented and discussed.

Instructing the judges how to review the test items.

- Review with the judges the purpose of the tests and the standard setting exercise including presentation and discussion of the district's definition of minimal competence. The definition may need elaboration or better description. The problems may become clearer once the items are under review. Be sure to make a distinction between average performance and minimal performance. Allow 3 to 4 hours to review an average length test (75 items).
- Distribute practice questions to judges.
- Describe the judges' task:

  The Nedelsky Method of standard setting requires a group of knowledgeable judges to inspect each question in a multiple-choice test and make a judgement about each wrong alternative. Each judge must decide whether a hypothetical student who just barely meets the district's definition of minimal competence could be expected to eliminate the wrong alternatives.

- After the task is clear to the judges, briefly discuss the rationale for eliminating some alternatives and not others. Pass out the test booklets to the judges. Begin discussion with the first question. Ask how many of the judges think that the first wrong alternative would be eliminated by the minimally competent student. How many disagree.

- For each wrong alternative that the group of judges does not unanimously agree upon, ask one judge from each viewpoint in turn to briefly explain his/her reasons. The purpose of such an exchange is not to force consensus but to allow different points of view to be heard from the group. Judges may change their minds after hearing the two or more different reasons. Once all of the wrong alternatives have been reviewed in this way, ask the judges to circle the number of alternatives they personally think that a minimally competent student should be able to eliminate as wrong answers.

- Go through each question on the test in the way described above. There will be a great deal of discussion at the beginning of the session. Later on, the reasons for disagreement generally will be the same as those already aired.

- During the review of the first several questions, circulate around the room to make sure the judges are circling their forms correctly.

- When all test items have been reviewed, ask the judges to tally up each column on their forms. Double check addition by summing together the column totals to make sure it equals the total number of questions, as shown on page 27.

- To calculate the standard, transfer the column totals to the corresponding probability and multiply. The sum of those multiplications is that judge's standard estimate. Average all of the judges' estimates to obtain the recommended group standard.

The judgements should be realistic. Hopefully the test has already been given to some defined set of students and item difficulties (p-values) are available from either the district's own data or a technical manual. Periodically monitor the judges as they raise their hands on each wrong alternative. A difficulty index can be quickly calculated for an item by using the probabilities on the recording forms. If the judges' expectations consistently exceed the average performance of a comparable population of students, the standard most likely will be unrealistically high. In such a case remind the judges of the purpose of the test and possible consequences (e.g., do they really think that 50% of the students do not meet minimum graduation requirements).

Example of forms filled out by three judges for a 10-item test

Judges Recording Form (Nedelsky Method)

Judge: #1

Question Number	Circle Number of Choices Identified			
1	0	1	(2)	3
2	0	(1)	2	3
3	0	(1)	2	3
4	0	1	2	(3)
5	(0)	1	2	3
6	0	(1)	2	3
7	0	1	2	(3)
8	0	1	(2)	3
9	0	1	2	(3)
10	0	1	2	(3)
TOTAL	1	3	2	4

(10)

Total 0's x 0.25 = 1 x .25 = .25
Total 1's x 0.33 = 3 x .33 = .99
Total 2's x 0.50 = 2 x .50 = 1.00
Total 3's x 1.00 = 4 x 1.00 = 4.00

SUM 6.24

1

Judge: #2

Question Number	Circle Number of Choices Identified			
1	0	1	2	(3)
2	0	(1)	2	3
3	0	1	(2)	3
4	0	1	2	(3)
5	0	(1)	2	3
6	0	1	(2)	3
7	0	1	(2)	3
8	0	1	(2)	3
9	0	1	2	(3)
10	0	1	(2)	3
TOTAL	0	2	5	3

(10)

Total 0's x 0.25 = 0 x .25 = 0
Total 1's x 0.33 = 2 x .33 = .66
Total 2's x 0.50 = 5 x .50 = 2.50
Total 3's x 1.00 = 3 x 1.00 = 3.00

SUM 6.16

Judge: #3

Question Number	Circle Number of Choices Identified			
1	0	1	(2)	3
2	0	1	(2)	3
3	0	(1)	2	3
4	0	1	2	(3)
5	(0)	1	2	3
6	0	1	(2)	3
7	0	1	(2)	3
8	0	(1)	2	3
9	0	1	(2)	3
10	0	1	2	(3)
TOTAL	1	2	5	2

(10) 31

Total 0's x 0.25 = 1 x .25 = .25
Total 1's x 0.33 = 2 x .33 = .66
Total 2's x 0.50 = 5 x .50 = 2.50
Total 3's x 1.00 = 2 x 1.00 = 2.00

SUM 5.41

To compute the cut-off score for this 10-item test, take the average of the three judges' SUMS.

$\frac{17.81}{3} = 5.94$

6.24
6.16
5.41
17.81

Sample Judge's Recording Form (Nedelsky Method)
for a 75 item, 4-alternative multiple choice test

Question Number	Circle Number of Choices Identified				Question Number	Circle Number of Choices Identified			
1	0	1	2	3	39	0	1	2	3
2	0	1	2	3	40	0	1	2	3
3	0	1	2	3	41	0	1	2	3
4	0	1	2	3	42	0	1	2	3
5	0	1	2	3	43	0	1	2	3
6	0	1	2	3	44	0	1	2	3
7	0	1	2	3	45	0	1	2	3
8	0	1	2	3	46	0	1	2	3
9	0	1	2	3	47	0	1	2	3
10	0	1	2	3	48	0	1	2	3
11	0	1	2	3	49	0	1	2	3
12	0	1	2	3	50	0	1	2	3
13	0	1	2	3	51	0	1	2	3
14	0	1	2	3	52	0	1	2	3
15	0	1	2	3	53	0	1	2	3
16	0	1	2	3	54	0	1	2	3
17	0	1	2	3	55	0	1	2	3
18	0	1	2	3	56	0	1	2	3
19	0	1	2	3	57	0	1	2	3
20	0	1	2	3	58	0	1	2	3
21	0	1	2	3	59	0	1	2	3
22	0	1	2	3	60	0	1	2	3
23	0	1	2	3	61	0	1	2	3
24	0	1	2	3	62	0	1	2	3
25	0	1	2	3	63	0	1	2	3
26	0	1	2	3	64	0	1	2	3
27	0	1	2	3	65	0	1	2	3
28	0	1	2	3	66	0	1	2	3
29	0	1	2	3	67	0	1	2	3
30	0	1	2	3	68	0	1	2	3
31	0	1	2	3	69	0	1	2	3
32	0	1	2	3	70	0	1	2	3
33	0	1	2	3	71	0	1	2	3
34	0	1	2	3	72	0	1	2	3
35	0	1	2	3	73	0	1	2	3
36	0	1	2	3	74	0	1	2	3
37	0	1	2	3	75	0	1	2	3
38	0	1	2	3					
Subtotal	—	—	—	—	Subtotal	—	—	—	—
					TOTAL	—	—	—	—

Total 0's x 0.25 =
Total 1's x 0.33 =
Total 2's x 0.50 =
Total 3's x 1.0 =
SUM

Nedelsky Method

Advantages	Disadvantages
1. Procedure is independent of numbers of students taking the test. Cut-off can be calculated for very small groups of students.	1. Procedure is blind to actual student performance on test.
2. Procedure can accommodate participation by a broad cross-section of community experts as judges (e.g., teachers, administrators, parents, students).	2. Cut-off can be too high or too low when the number of judges with the same special interest or bias are disproportionately represented.
3. Procedure is based on close scrutiny of items in the instrument.	3. When different forms or tests are used that are not equated, the procedure must be repeated for each form or test.
4. Procedure most closely follows decision processes of test takers, each alternative for each item must be rejected or accepted.	4. More time and people are needed to make judgments than with performance based procedures.*
	5. Can only be used with multiple-choice test.

* Unless one counts the test administration, scoring, etc., involved in the contrasting groups approach.

Contrasting Groups Method (Method A)

The Contrasting Groups Method requires judgements about the students who will take the test rather than judgements about the instrument itself.

The method assumes that there are three distinguishable groups of students in the school population. These groups were described at the beginning of this chapter as "non-masters", "borderline" and "masters" of the district-defined basic skills. The best a priori and independent judgement of which student can be best classified in which group can be made by a teacher who has been teaching the student in the skill area long enough to be reasonably certain of that student's achievement.

Several conditions are necessary to obtain valid teacher judgements.

- Judgements must be made on the basis of the student's achievement in the basic skills not on the student's attitude, attendance or other behavioral problems. A student's mastery of a basic skill often may be difficult to isolate from these variables. Scores from such students, as well as recent transfers, should not be used in calculating the standard.

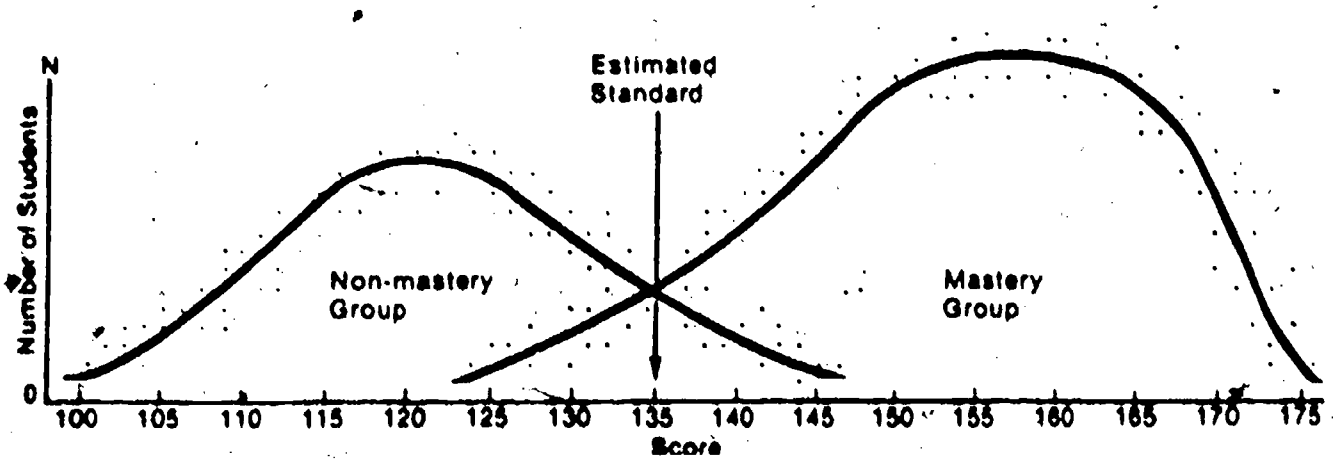
- The judgements should be made by teachers who know the student's skills with respect to the test content. For example, math teachers should judge students for the math test and English teachers should judge reading and writing. Depending on the course structure and content, others such as social studies teachers may be able to make perfectly adequate classifications. Even math and English teachers who are teaching higher level courses (e.g., algebra and literature) may have difficulty focusing their judgements on the basic level of mastery.

- The district's definition of mastery or minimal competence in the skill area should be clear to the teachers. The purposes of the test and reason for collecting their judgement should also be clear.

- Judgements must be based on the students' present achievement. Judgements about a student's performance that are even a few months old may not be accurate indicators of what the student has retained or has learned subsequently.

- Judgements must be made independent of the test scores, i.e., without knowledge of how the students actually performed on the proficiency test.

The standard obtained with the Contrasting Groups Method is based on the test performance of two groups of students identified by knowledgeable judges. Judgements made by teachers independent from the test scores classify the students into those who are clearly masters of the skill and those who are clearly not masters of the skill. Scores from students who cannot be classified into either of those groups are not used in calculating the standard. The standard is placed generally at the point where the two distributions of scores for "masters" and "non-masters" intersect as shown below. This is the point where both types of classification errors can be held to a minimum as discussed earlier in this chapter.



If possible, there should be at least 100 students in the district who would be classified as "non-masters." If fewer than 100 students are in the above or below standard groups, the standard proposed by the

intersection of the two distributions should be viewed with extreme caution. Some attention should be given to the means and standard deviations of these two groups, and more specifically to how a score at the first standard deviation above the mean for the below standard group compares with a score at the first standard deviation below the mean for the above standard group.

Procedurally the steps for collecting teacher judgements are as follows:

- Identify the students and classrooms for which appropriate judgements can be made.
- Obtain the class roster.
- Have teachers who can make valid judgements about those students for each skill area meet to explain procedures and become familiar with the definition of what is clearly a "master" and a "non-master" of the district's objectives.
- Have teachers make their judgements of the students they know and mark their rosters, for example, with a "1" next to those students who are clearly masters and a "3" next to those who clearly need supplemental instruction. Any student who is too close to the borderline or who presents problems in classification because of behavior, attendance, or who is not known by the teacher should be marked with a "2".

After the judgements have been collected and the test administered, enter the test score data and the judgements on the data entry sheet.

Contrasting Groups Method

Advantages	Disadvantages
1. Procedure is based on actual student performance data.	1. For reliable calculations, should have about 100 students in smallest of two groups.
2. Judgments based on teachers, who are knowledgeable about subject matter and students.	2. No direct input by affected constituencies except teachers of students being tested.
3. Psychometrically cleaner - errors of classification are minimized and can be identified by case.	3. May not be able to obtain judgments for students who are not taking a course in the subject-matter being tested.
4. Does not require as much time to collect judgments as item-review procedures.	4. Must test later in the year than with item-review procedures to allow teachers time to know students' abilities.
	5. If definitions of mastery and non-mastery are not clear and uniformly applied by teachers there is a danger of misidentification. Teacher judgments may not be made with respect to test or district's definition of minimal competence.

Standard Setting on Writing Samples

Proposed standards on the writing sample will be developed by the application of the contrasting groups method to the students' writing scores and the judgements of scores reviewed by the grade level committee. A packet of scored essays with the scores removed will be provided to the appropriate grade level committee to allow them to identify above and below standard papers in a way that would identify their proposed standard. This same packet will be made available to the advisory committee in its review of the proposed standards.

Note: The contents of this section were adapted from several working documents prepared by Educational Testing Service including the Manual for Setting Standards prepared by M. Zieky and S. Livingston.

Data Collection Procedures

The data collection procedures in this study will result in various products which the local district will develop. The following is a list of the products to be developed along with some comments on how they can be developed.

Products	Developed by
1) Lists of basic skill objectives in reading, writing and mathematics at three grade levels	grade level committees (or other curriculum committees already established)
2) Initial list of tests to be considered at each of three grade levels	the facilitator, working with grade level committees
3) Completed test review forms on one or more tests at each of the three grade levels	the grade level committees (the facilitator then orders the tests selected and arranges for the administration of the tests)
4) Completed item-by-item review for standard setting for each objective test selected at each grade level	the grade level committees
5) Completed writing samples at specific grade levels (grade 8-9 at least)	administered by classroom teachers and coordinated by facilitator
6) Writing samples coded (system, school, teacher codes)	clerical assistant to facilitator (codes kept "confidential")
7) Roster lists of student scores on writing sample	classroom teachers
8) Standardized test scores - reading, mathematics and, at some grade levels, possibly writing (or language arts)	test publisher or scoring service, following local administration of the tests
9) Teacher ratings of all regular students at the three grade levels, identifying "masters" and "non masters" in reading writing, and mathematics	a survey form (see Appendix G) completed by all teachers in these subjects at the three grade levels. (survey administration to be coordinated by the facilitator)

10) Completed Data Entry Sheets (See Appendix H) for all regular students at each of the three grade levels	clerical assistant to facilitator

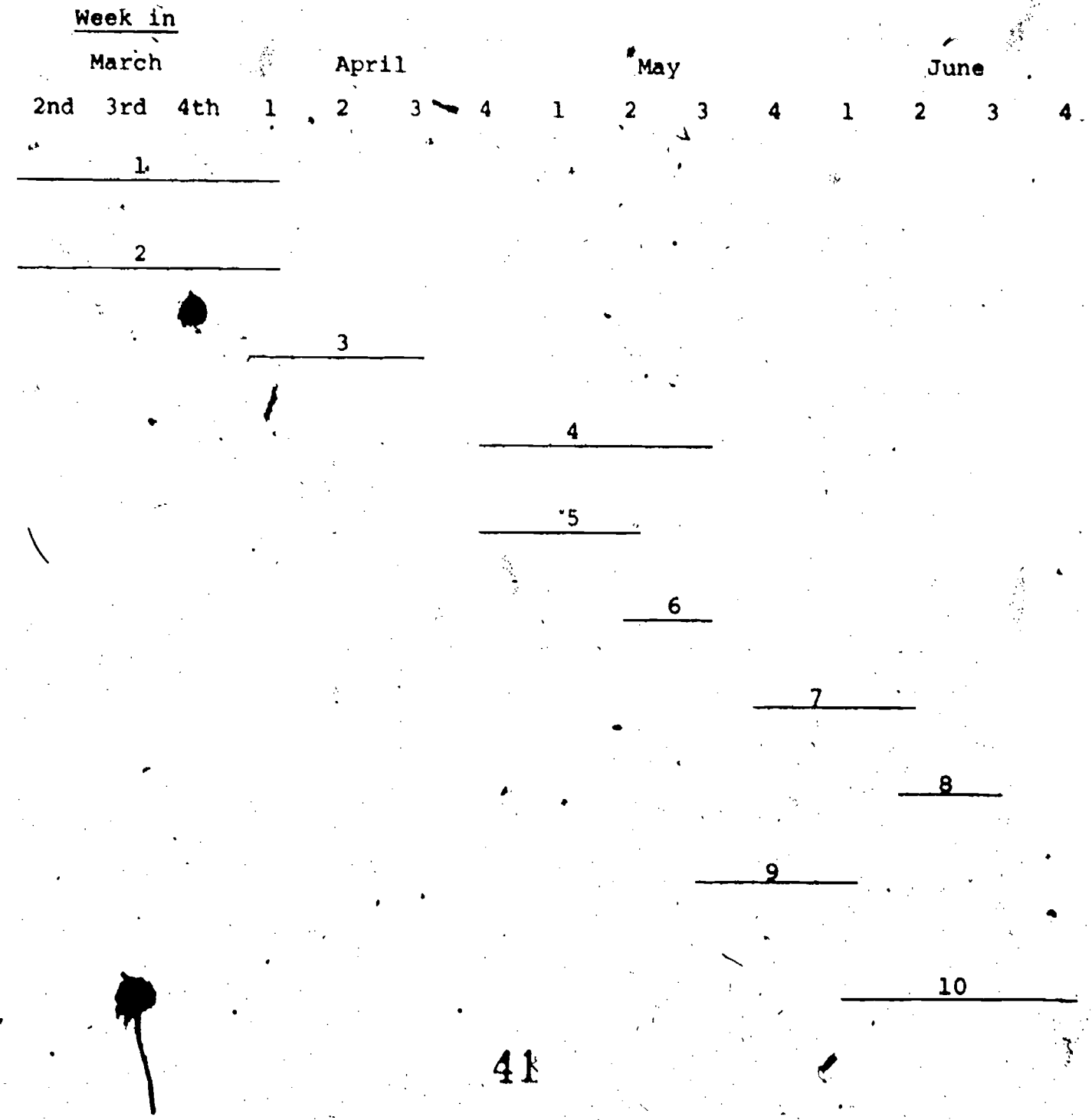
The first nine products (#1-9) make possible the completion of product #10, the data entry sheet for all of the regular students. On these data entry forms, the following information is to be recorded for each student:

- ° student code (names need not be included; space is available for the name only for the convenience of the local staff)
- ° grade level
- ° sex
- ° minority group status
- ° status in reference to English or non-English speaking
- ° most recent marks in reading, writing, mathematics
- ° test or writing sample scores in all three subjects
- ° teacher ratings in all three subjects

For each grade level, a cover sheet will be required providing information on the score data included on the data entry form (range, intervals, etc.)

Figure 1 outlines a tentative schedule for the collection of data, assuming that the local effort will get underway in early March with testing in early May. Districts conducting their testing in the Fall would follow a somewhat similar schedule. These districts would collect data for the writing standards according to the March-June schedule shown below. In the Fall, the other teacher ratings would be delayed until November, allowing the teacher sufficient time to work with the students before rating their needs for special work in basic skill development.

Figure 1: Tentative Schedule for Data Collection Activities



Most of the tasks and resultant products listed above need little or no comment in addition to what has been presented in earlier sections. Two data collection tasks, the teacher rating survey and the completion of the data entry form do merit additional comment.

The teacher rating survey will require a matching of students with those teachers who have worked with them in reading or writing or mathematics. In self contained classrooms, the one teacher would rate each student in all three skill areas. In a departmental or team arrangement, the circulation of copies of the class roster (one for each teacher), with space for ratings in all three subjects (See Appendix G) would enable each teacher to submit ratings in the appropriate subject areas without being influenced by the ratings assigned to these same students by other teachers. A less desirable alternative would be the circulation of one class roster list with teachers indicating all three ratings on the same list.

The final step in the data collection process is the completion of the data entry form described above. These data entry sheets could be filled in gradually with data entered as they become available. The school marks used in this study could be those from the previous semester, if the use of current semester marks would unduly delay the process. These marks are included simply to give some additional information on how a proposed standard affects "A" students, "B" students, "C" students, etc. The assumption here is that, on the whole, marks do not fluctuate substantially from one semester to the next, despite individual ups and downs. In completing the data entry form, it is essential that the

format be followed exactly. The data are key-punched directly from these forms into a standard computer program. The completed data forms will be collected by ETS in the last week of June, from those districts on the March-June schedule.

Report Interpretation

In some ways, the new state requirements are not all that new. Basic skills improvement has been a primary goal of schooling from the earliest days. With the Elementary and Secondary School Act of 1965, formal programs such as Title I programs, for students in need of special assistance, have been much in evidence in most schools. For many of these programs, particular levels of attainment on various tests have been used to guide decisions on student placement into these programs. While the present basic skills improvement policy requires, in effect,a similar screening for placement purposes, the context of this new requirement adds a note of accountability to local standard setting. For this reason, more formal attention to standard setting is called for. With this new requirement, the local standard publicly defines the expectations of the local school system, in ways that Title I entry level requirements never did. For these reasons, the school staff and community representatives must work together to establish the performance standards.

One major task of the project facilitator is to work with the system-wide advisory committee, using the MEC-ETS standard setting study report, to formulate performance standards to be proposed to the school committee. A draft copy of the report, with data from a fictitious school district, is included in Appendix A. Section I of that report proposes two possible standards in each of the three skill areas at each of the three grade levels and then discusses the impact of each of those proposed standards. Section V of the report presents more detailed analysis of teacher ratings, test scores, and school marks in each skill area at each grade level, along with information on the impact of proposed standards on students by sex, race, and language status.

The standards to be set are indeed local standards, proposed by the advisory committee and established finally by the school committee. The two standards proposed in the MEC-ETS report are simply recommendations based on the data presented.

In reviewing these recommendations, the advisory committee might find the following questions useful for purposes of discussion.

- o What is the impact of each proposed standard in terms of the percentage of students falling below? Is this reasonable for this community? (For example, a community which perceived itself as having a serious problem with basic skill development might find a standard at the 25th percentile of local performance more acceptable than one at the 5th percentile.)
- o What is the impact of each proposed standard on students who received "A's" last semester? "B's"? "C's"? etc.
- o What kind of remedial program would prove feasible given present and anticipated resources? Which of the proposed standards is most feasible in this regard?
- o Should one standard be adopted for the present and a higher ("goal") standard scheduled to go into effect in three to four years?
- o How should the local standards relate to the national or state "average" (50th percentile)?

Standard setting is in the final analysis a matter of judgement. Informed judgement, however, is always to be prefereed to casual judgement. The approach taken here allows the local committee to approach the decision with as much information as reasonably can be brought to bear on the decision. Whatever decision the committee makes, the standard recommended and

finally established should relate to some pertinent information, going beyond the immediate appeal of certain numbers (such as 70 percent or 80 percent correct).

The state requires the reporting of the number of students by sex, race, and language background who fall below the standard. All of the procedures used in the MEC-ETS standard-setting study are based on total group, not sub-group data. No bias for or against any sub-group has been introduced by these methods. Some might argue that some possibility of bias is inherent in mathematics and in the reading and writing of standard English. Any consideration of bias must distinguish between the possible bias in the requirement itself and bias in the manner in which the requirement is met. In this latter regard, there are two areas in which local districts should make certain they are free from any bias. One is in the selection of the tests used and the other in the setting of the standard.

All tests used should be screened for sex, racial, or other kinds of bias by a careful review of the contents of the tests themselves.[1] (For additional information, see Appendix K.)

The standards established by the impartial procedures described above should be fair and unbiased. In reviewing the options provided by these procedures, the local advisory committee should take into account the impact on various sub-groups of any proposed standards. This should not be construed as a recommendation that standards necessarily be lowered to accommodate any sub-group. Because the standard

[1] Where needed, item analysis data by sub-group might be examined to show the test "works" the same way for each sub-group.

simply defines the number of students who will receive special attention in the development of their basic skills, higher standards are not necessarily disadvantageous to a group that has difficulty with the requirement. In planning subsequent learning activities, attention must be given to any segregative effect a particular program might inadvertently introduce.

Appendix A

Draft of MEC-ETS Standard Setting Study Report
to the _____ Advisory Committee

Introduction - Why this kind of study?

The _____ schools have had standards for years. These standards have been embodied in day to day teacher judgements, in schools marks, and in promotion/retention policies, to name but several of the ways in which schools maintain standards. The January 1979 decree of the State Board required each district to set explicit performance standards on basic skills tests at three grade levels. As used in this context, a performance standard is the minimum acceptable score on a basic skills test. While the implicit standards have been in effect for years, the setting of explicit performance standards is a new activity.

As school districts approach this new activity, they can choose between two possible approaches to setting standards. One approach assumes that in this business of setting standards there is "really nothing to it". The other approach calls for a standard setting study.

The "really nothing to it: approach involves a committee (or an individual) simply selecting a performance standard that they think looks reasonable. For example, they might set as a standard, 80 percent correct on any basic skills test they use. If they are using a test with local norms they might select the 30th percentile (12th grade) for their standard. The standards set

by this rather casual process could turn out to be quite fair and appropriate. They could also turn out to be quite unfair to the students' and school's reputation. This would happen, for example, if the tests used were so difficult that 50% correct would really prove to be the reasonable standard. In some schools, students below the 30th percentile could be well above minimum standards; in other schools, students at the 45th percentile might be well below acceptable standards. While the "really nothing to it" approach can, at times, work well, it can also work out badly. The chief problem with this approach is that the standards set by this process are almost impossible to defend in the face of any kind of challenge. Casual judgement is difficult to defend.

The _____ school committee has chosen the standard setting study approach which bases standard setting on information about the minimally qualified student and the difficulty level of the tests on which the standards are established. This approach also takes into account the impact of these performance standards in terms of the numbers of students above and below the standard, the relationship between these standards and school marks, and in a general way, the relationship between these standards and those being considered by neighboring districts. This information will provide the basis for _____'s decision in setting standards. Indeed, all standard setting is a matter of judgement. According to the State Board policy, the judgement involved in the state required standard setting in basic skills must be local judgement,

ultimately the judgement of the school committee. This local judgement will be advantaged by well organized, relevant information provided by this standard setting study. Judgement based on careful study can be readily defended.*

_____ Standard Setting Study

This report is designed to provide information to be used by the _____ advisory committee in formulating the specific performance standards it will propose to the _____ School Committee. This report presents proposed performance standards in reading, writing, and mathematics at the three grade levels, grades __, __, and __, selected by the _____ School Committee. This report includes the following five sections:

1) proposed performance standards for _____
2) an overview of the 1979 state requirements
3) the rationale for the tests used in this study
4) the rationale for the standard-setting procedures [redacted] in this study
5) detailed information in support of the proposed standards

This report has been prepared by the Merrimack Education Center (MEC) with assistance from the New England Office of Educational Testing Service (ETS) for use by the district-wide advisory committee in

---

* For example, the January 1978 decision of the U.S. Supreme Court upheld the teacher certification standards set on the National Teacher Examinations by the state of South Carolina.

drafting the standards it will propose for the _____ School Committee. In an appendix to this report, two sample report formats are included, a format for the advisory committee's report to the school committee and a format for the final report of the _____ School Committee in which the committee formally sets specific performance standards. These report formats have been developed simply to expedite the work of these committees.

SECTION I

Proposed Performance Standards

This section of the report presents the proposed performance standards for _____ in reading, writing, and mathematics for grades __, __, and __ based on the standard-setting studies conducted in cooperation with MEC and ETS. (A description and rationale for this study is presented in Section 4. The selection of the tests to which the standards are related is described in Section 3.) In presenting the proposed standards, the following information will be provided for each standard.

- the standard (a score on a given test or exercise)
- the method used in arriving at this standard
- the national (or state) percentile equivalent (where available)
- the number of and percent of students at or above the standard
- the number and percent of A, B, C, and D students at or above the standard (where available)
- the relationship between the proposed standard and corresponding standards proposed by MEC member districts (where available)

comments on classification and mis-classification issues

To preserve the complete autonomy of the _____ committee, this MEC-ETS report will not go beyond the presentation of the relevant information described above.

Where possible, two standards are proposed for each test at each grade level, a Method A and a Method B standard. In this study, teachers were asked to identify those students who definitely were in need of remediation in basic skills in mathematics (below standard) and those students who were definitely not in need of such assistance (above standard). A comparison between the scores of these two contrasting groups on the _____ test produced the standard reported here as the Method A standard. Teachers were also asked to review the items in the test and estimate the performance on those items of minimally qualified students. This produced the standard reported here as the Method B standard.

Proposed Standards - Grade 3 , Reading

At grade __, the _____ test was selected as an appropriate measure of basic skills in reading . On this test (Level __, Form ___), the Method A proposed standard is a scaled score of which is at the ___ile of national performance (grade ___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in reading , ___ percent were above the standard, as were __percent of the B students, ____ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the correspohding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and ___ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___%ile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in reading , ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 3 , Writing

At grade__, the_____test was selected as an appropriate measure of basic skills in writing . On this test (Level__, Form___), the Method A proposed standard is a scaled score of , which is at the ___tile of national performance (grade___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in reading , ___ percent were above the standard, as were __percent of the B students, ____ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is [redacted] the ___tile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in writing , ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 3 , Mathematics

At grade__, the_____test was selected as an appropriate measure of basic skills in mathematics. On this test (Level__, Form___), the Method A proposed standard is a scaled score of , which is at the ___%ile of national performance (grade___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in mathematics ,___ percent were above the standard, as were __percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___ile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in mathematics, ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method ___ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 6 , Reading

At grade__, the_____test was selected as an appropriate measure of basic skills in reading . On this test (Level__, Form___), the Method A proposed standard is a scaled score of , which is at the __%ile of national performance (grade___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in reading ,___ percent were above the standard, as were __percent of the B students, ____ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___%ile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in reading , ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a <u>rough basis</u> for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent <u>more</u> students to be classified at or above standard.

Proposed Standards - Grade 6 , Writing

At grade __, the _____ test was selected as an apprppriate measure of basic skills in writing . On this test (Level __, Form___), the Method A proposed standard is a scaled score of , which is at the ___tile of national performance (grade___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in writing , ___ percent were above the standard, as were __percent of the B students, ___ percent of the C students and ___percent of the D students, In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___tile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in writing , ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 6, Mathematics

At grade __, the _____ test was selected as an appropriate measure of basic skills in mathematics. On this test (Level __, Form ___), the Method A proposed standard is a scaled score of , which is at the ___%ile of national performance (grade ___, Spring). At grade ___, ___ students or __ percent of the total number were at or above this standard. Among A students in mathematics , ___ percent were above the standard, as were __percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___ile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in mathematics, ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 9 , Reading

At grade __, the _____ test was selected as an appropriate measure of basic skills in reading . On this test (Level __, Form___), the Method A proposed standard is a scaled score of , which is at the ___ile of national performance (grade ___, Spring). At grade ___, ____ students or __ percent of the total number were at or above this standard. Among A students in reading , ___ percent were above the standard, as were __percent of the B students, ____ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___ile of national performance (grade ___, Spring). At grade ___; ___ students or ___percent of the total number were at or above this standard. Among A students in reading , ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method __ standard allows ___percent more students to be classified at or above standard.

Proposed Standards - Grade 9 , Mathematics

At grade __, the _____ test was selected as an appropriate measure of basic skills in mathematics. On this test (Level __, Form __); the Method A proposed standard is a scaled score of , which is at the ___%ile of national performance (grade ___, Spring). At grade __, ____ students or __ percent of the total number were at or above this standard. Among A students in mathematics , __ percent were above the standard, as were __percent of the B students, ____ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used as a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In reviewing the effectiveness of Method A in this instance, it should be noted that ___ percent of the students were misclassified as masters and __ percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

The Method B proposed standard is a scaled score of ___, which is at the ___ile of national performance (grade ___, Spring). At grade ___, ___ students or ___percent of the total number were at or above this standard. Among A students in mathematics, ___percent were above the standard as were ___ percent of the B students, ___ percent of the C students and ___percent of the D students. In comparison with standards proposed by the majority of the other MEC districts, this standard fell (within, above, or below) the range of commonly proposed. In comparison to the total range of standards proposed, this standard fell within, above, or below the range. In these comparisons the corresponding national percentile scores were used on a rough basis for comparison. These percentiles describe the performance of separate but allegedly similar national norm groups on different basic skills tests.

In comparing the Method B standard to the Method A standard, the Method ___ standard allows ___ percent more students to be classified at or above standard.

Performance Standards - Grade 9, Writing

The state regulations require a writing sample as the measure of basic skills in writing. The scoring of writing samples call for a different set of procedures than the scoring of objective, multilple choice tests. These procedures call for at least two readings of each paper by readers using the same scale of scores. This uniformity in scoring is obtained through a training exercise conducted immediately before the scoring session. MEC with assistance from ETS conducted a training and scoring session May 9. During this session, papers from all MEC districts using this standard setting study service were scored by some thirty five teachers from all participating distriots. Papers were coded so that no district or school could be identified by any of the scorers.

At the conclusion of this scoring session, the teachers from each district were asked to select a score that represented the minimum performance standard - the minimum level of acceptable work. The _____ teachers selected the score of ___ for the _____ standard. This score was reviewed and approved by a committee of eighth and ninth grade _____ teachers, after they reviewed the range of papers at various score points. This standard is, in effect, the Method B standard on this exercise. Applying this standard to the papers scored on May 9, ___percent of the _____ students were at or above the standard.

The Method A standard applied to this same group of students showed ___percent at or above the standard. In reviewing the effectiveness of Method A in this instance, it should be noted that ___percent of the students were mis-classified as masters and ___percent as non-masters. The proposed Method A standard does, however, fall between one standard deviation below the mean of the above standard group and one standard deviation above the mean of the below standard group, indicating that the more typical masters and non-masters are accurately classified.

In comparing the Method B standard to the Method A standard, the Method ___ standard allows ___percent more students to be classified at or above standard.

SECTION II

Overview of 1979 State Requirements on Basic Skills Improvement

This section will present an overview of the new state regulations similar to the overview presented in Section 1 of this manual

SECTION III

Rationale for Tests Used in the Study

This section will review the procedures followed and information developed in the process of test selection as described in section 4 of this manual. To complete this section of the report, the facilitators would need to make available for each test used the following:

1) List of basic skill objectives to be covered by the test.

2) Completed Item review form

3) Information on score reports available

4) Information on test reliability and other statistical information, when appropriate

SECTION IV

Rationale for Standard-Setting Procedures Used

This section will describe the reasons for selecting the two approaches to standard-setting used in this project. Justification will be presented for the reliance on professional judgement at the outset, followed by a review and confirmation (or modification) of that judgement by community representatives. Material for this section will be drawn to some extent from Section 6 of this manual.

SECTION V

Detailed Information Related to Proposed Standards

This section will present supportive data for the standards proposed in Section I of this report. These data will include a set of tables on each test, a summary of rating scores on all objective tests related to the Method B standard.

In each subject/skill area tested, Table A will present the teacher ratings for all those grade levels. Table A is presented here with fictional data.

Table A: Teacher Ratings of Students in Math, Grade 3, 5 & 9.

Grades	3		6		9	
Teacher Ratings	N	%	N	%	N	%
Above Standard	200	40	240	48	340	68
Possibly at Standard	150	30	160	32	100	20
Below Standard	150	30	100	20	60	12
Total N	500		500		500	

Tables B and C will present the scores related to the teacher ratings in each subject at each of the three grade levels. The Method B standard will also be linked to these two tables so that the reader can see the impact of each standard both on the total group and on students grouped by school marks. These same three tables will be presented by race and sex for each skill area at each grade level to provide the advisory committee with a complete picture of the impact of the proposed standards.

Tables B and C are presented here with fictional data.

Table B: Math Test Scores of Above and Below Standard Students, Grade 6

Test Scores (Scaled Scores)	Above Standard N	%	Below Standard N	%	Total Group N	%
199-90	20	8			20	4
189-80	40	16			65	13
179-70	70	30			110	22
169-60	50	20			110	22
159-50	30	12			55	11
149-40	20	8	10	10	40	8
139-30	10	4 *	30	30 *	40	8 *
129-20			40	40	40	8
119-10			10	10	10	5
109-00			10	10	10	5
	240		100		500	
Mean.....	169.6		125.2		160.5	
S.D.....	15.30		9.65		21.3	

Proposed Standard..... 137

Table C: Math Test Scores and Grades (Previous Semester), Grade 6

Grade	A (5)		B (4)		C (3)		D (2)		F (1)	
Test Scores (Scaled Scores)	N	%	N	%	N	%	N	%	N	%
199-190	15	13	5	3						
189-180	30	28	25	16	10	6				
179-170	20	18	65	41	25	14				
169-160	40	37	40	25	30	17				
159-150	5	4	10	6	30	17				
149-140			10	6	20	12	10	25		
139-130			5	3	25	14	10	25		
129-120					30	17	10	25		
119-110					5	3	5	12		
109-100							5	12	5	100
Total N	110		160		175		40		5	
Mean	170.1		165		142.9		123.7		105.1	
S.D.	11.6		12.8		24.5		13.3		1.2	

The comments will be made to each set of tables calling the attention of the advisory committee to the salient information presented in each table.

(While this list deals chiefly with high school level tests, the general batteries include tests of the lower levels as well.)

Currently Available Tests

A number of tests are currently available for assessing basic skill development at the high school level. This number includes specialized test programs, designed to meet a rather specific set of needs, and the general achievement test series which are designed for large volume use in systemwide testing programs. The general achievement tests are listed in Table 5.1.The following are some specialized tests described briefly in alphabetical order.

ABLE, the Adult Basic Learning Experiment, developed by the Psychological Corporation, tests at a high school level basic vocabulary, reading, spelling, and arithmetic used in everyday adult life. The test battery, which includes 198 items, takes three and one half hours to administer.

APL, Adult Performance Level Survey, developed in a University of Texas based research project and published by the American College Testing Program (ACT), tests at a high school level basic skills in reading, writing, computation and problem solving in the areas of consumer economics, occupational knowledge, community resources, health, and government and law. This 40 item survey takes approximately one hour to administer.

BSAP, the Basic Skills Assessment Program, developed by Educational Testing Service in conjunction with a nationwide consortium of school districts, tests, at the junior high and high school level, basic competencies in reading, mathematics, and writing skills. These three tests, which together include 220 items, take approximately two hours to administer. This program is unique in that it provides an ongoing secure test program with procedures for local standard setting.

EDST, the Everyday Skills Test, developed by CTB/McGraw Hill, assesses student mastery of reading and math skills at the intermediate level (grades 6-12) using items with a practical orientation (job applications, discount rates, etc.). These tests require approximately 2-1/2 hours to administer.

GED, the General Education Diploma Test, published by the American Council on Education, is the commonly used high school equivalency test for adults. These tests cover English expression, Reading in Social Studies, Science, and Literature, and general Math. Spanish and French language versions are available. The use of these materials in high schools would require special negotiations.

IOX Basic Skills Tests, developed by the Instructional Objectives Exchange, tests skills in reading, writing, and mathematics using "real life" content. Two 45-minute forms are available for each of the three tests.

SHARP, the Senior High School Assessment of Reading Performance, developed by CTB/McGraw Hill in cooperation with the Los Angeles Unified School District, tests minimum competencies in reading skills deemed necessary for everyday life. SHARP consists of thirty displays of everyday reading materials, matched to behavioral objectives. A parallel computational

skills test (TOPICS) is scheduled for publication in February, 1979. A Writing Skilsl Test and Writing Sample will be available in the Fall of 1978.

SRA Survival Skills Tests, published by the Science Research Associates, test reading and mathematic skills on everyday, practical items (sales slips, recipes, road signs, caution labels) in a school edition (grades 6-12) and an adult edition (grades 11, 12, and above). This 120 item battery requires approximately 2 hours to complete.

TABE, the Test of Adult Basic Education, published by CTB/McGraw Hill, tests basic reading, mathematics and language skills at the junior high school level. This newly published shorted version of the California Achievement Test listed below includes 328 items and requires 3 hours to administer.

Services for Local Test Development

In addition to published materials, several test publishers provide services that enable districts to have tests developed to match objectives which the districts select from a bank of objectives developed by the publisher. The SCORE service, for example, offered by Houghton-Mifflin, enables a distrfct to choose from a bank of 1700 objectives (K to Grade 9) in reading, language arts, and mathematics. Test copies are prices for orders of 1,000 or more on the basis of the number of pages in the test (e.g., 1000 copies of an 8 page test would cost .98¢ a copy). CTB/McGraw Hill offers a similar service called ORBIT with over 800 (K to Grade 12) objectives in reading, communication skills, and mathematics. The Instructional Objectives Exchange offers a similar set of services. SRA's mastery program would provide customized test services in the area of basic skills assessment.

Table 5.1 GENERAL ACHIEVEMENT TESTS COVERING BASIC SKILLS AT THE HIGH SCHOOL LEVEL

TEST	LEVEL	GRADE	CONTENT	PUBLISHER
California Achievement Test	4-5	6-12	Reading Mathematics Language	CTB/McGraw Hill
Comprehensive Test of Basic Skills	3-4	6-12	Reading Language Math Reference Skills Science Social Science	CTB/McGraw Hill
Iowa Tests of Basic Skills	7	9-14	Vocabulary Reading Comprehension Language Skills Work Study Skills Math Skills TOTAL	Houghton-Mifflin
Iowa Tests of Educational Development		9-12	Reading Language Arts Mathematics Social Studies Science Use of Sources	Science Research Associates
Sequential Tests of Educational Progress	2 & 3	7-12	Reading Writing English Expression Math Computation Math Basic Concepts Science Social Studies	ETS/Addison-Wesley
Stanford Test of Academic Skills	I & II	8-13	Reading English Math	Psychological Corp./HBJ
Tests of Academic Progress	Form S	9-12	Social Studies Composition Science Reading Math Literature	Houghton-Mifflin

Test (Level, Form) _____

Basic Skills Item Review Form

Grade _____ Page _____

Item Number	Content Taught			Skill Level		Related to Local Objective (cite #)	Comment
	Earlier	This Year	Later	Basic	Advanced		

HOLISTIC SCORING

I. What is holistic scoring?

A. The theory.

1. The whole of a piece of writing is greater than any of its parts.

2. English teachers, though they may have difficulty in giving a verbal description of writing ability that is recognizable to all, can recognize good writing when they see it.

3. Though in an analytic reading teachers may not agree on the weight to be given a particular trait, these same teachers will, in judging a work as a whole, rank papers in much the same way.

4. No aspect of writing skill can really be judged independently; the halo effect is always strong.

B. The method.

1. The standards.

a. Standards are not imposed upon readers; readers themselves determine standards.

b. Papers are not judged against an ideal, but against what is: what students have written on this topic at this time.

c. Standards must be maintained and reinforced throughout the reading.

2. The judgments.

a. Judgments are made on anonymous papers.

b. Judgments are independent.

c. Multiple judgments on each paper are mandatory.

d. Judgments must be quick and immediate.

e. Judgments must be definite, for the score scale has no middle points.

3. The scoring.

a. The score is the sum of all the readers' judgments.

b. Some discrepancies in the scores the readers give are to be expected.

c. Wide discrepancies between readers' scores must be corrected immediately.

d. Regular divergence from the standards on the part of any reader must be corrected.

II. Why use holistic scoring?

A. It is efficient.

B. It is reliable.

C. It emphasizes what is right rather than what is wrong with a piece of writing.

D. It requires consensus among readers.

E. It encourages evaluation of the program, as well as the individual pieces of writing.

III. How is a topic scored? (Actual reading)

A. The topic is read and analyzed.

B. The ground rules are established.

C. The standards are set through the use of sample papers.

D. The papers are read.

1. First reader's score must remain unknown to other readers.

2. All papers should be read once before any are read twice.

3. Readers must be allowed to rest regularly.

4. Papers must flow efficiently from reader to reader.

IV. What makes a good topic?

A. The interest to the students.

B. The interest to the readers.

C. The range of writing it produces.

D. The relative objectivity with which it can be scored.

V. Of what use is holistic scoring in the schools?

A. It can promote communication about the teaching of writing among faculty members.

B. It can be used to measure growth in students' writing ability.

C. It enables teachers to score writing assignments quickly and reliably.

D. It calls for multiple evaluations.

TOPIC

HOLISTIC ESSAY SCORING WORKSHOP

If we listen to its critics, television is to blame for half the things wrong with our lives -- everything from our poor reading habits to the high crime rate. Very few people defend television, and yet it must serve some worthwhile purpose.

What values do you see in television -- to an individual, to a family, to society? Discuss one or two of these values, telling what each value is and how it benefits people. Use specific examples to support your ideas.

- Twenty Minutes -

TO BE ASSIGNED TO APPROXIMATELY FIFTY STUDENTS IN A CLASSROOM SITUATION

EDUCATIONAL TESTING SERVICE MIDWESTERN REGIONAL OFFICE

960 Grove Street
Evanston, Illinois 60201
312-869-7700

Memorandum for: Holistic Essay Scoring Workshop Participants

Subject:- Scoring Codes for the workshop essays

Enclosed please find the essays submitted from your class(es) for the workshop instruction.

As you will recall, each workshop participant indicated his or her score for an essay by using a letter code. That is, instead of recording the score 4, 3, 2, or 1 on the essay paper, each reader indicated the appropriate score by an assigned letter. Each reader had a unique code; however, in all codes assigned, a particular letter always represented the same value. For example, whenever B was assigned, it represented the value of 4; whenever G was assigned, it represented the value of 1. The basic code follows, indicating the letter equivalents for the four scores:

(high)	4.	B	E	F	K	P	R
	3.	I	J	L	S	T	Z
	2.	M	N	V	W	X	Y
(low)	1.	A	C	G	H	O	Q

Each paper was read at least once. About half the papers were read at least twice by two different readers. Of those papers scored twice, many were checked for discrepancies. The scores given by both readers were accepted if: 1) the same value was assigned by both readers to the essay (i.e. 4-4, 3-3, 2-2, 1-1) or, 2) the two values assigned by the readers were within one point of each other (i.e. 4-3, 3-2, 2-1). On the other hand, papers with a discrepancy of more than one point between readers (i.e. 4-2, 4-1, 3-1) necessitated a third reading. Note again though, that not all the papers were read twice and of those papers read twice, not all were checked for discrepancies. Therefore, all papers truly necessitating a third reading were not read a third time due to time constraints.

By noting the above code, you will be able to determine the values assigned to each essay written in your class(es). Bear in mind, the standards were set by reading papers from all grade levels and from different schools. While decoding the scores will facilitate feedback to your students, caution should be taken in interpreting these scores.

Again, it was a pleasure working with you. Please feel free to call upon us if we can be of assistance in any way.

MEC IMPROVING BASIC SKILLS PROJECT

TEACHER RATING FORM

Directions for Teachers

The new Massachusetts Basic Skills Improvement Policy requires local school districts to establish minimum performance standards in basic skills in order that additional attention may be given to students who are below this minimum standard.

Using your class list, you are asked to rate each student whom you are now teaching reading or writing or mathematics. Basic skills do not necessarily describe the full range of skills taught at any one grade level. A sample listing of basic skills at the 9th grade level is attached. In each skill area you now teach, please rate each student in reference to basic skill mastery at his or her particular grade level, using one of the following ratings:

"1" This student is below minimum standards and definitely needs remedial help

"2" This student is close to minimum standards and may or may not need remedial help

"3" This student is above minimum standards and definitely does not need remedial help

THESE RATINGS WILL BE COMPLETELY CONFIDENTIAL. THEY WILL BE USED ONLY FOR GROUP DATA ANALYSIS AND WILL NOT AFFECT THE STUDENT IN ANY WAY.

Appendix G

CLASS LIST

SCHOOL_____ GRADE LEVEL_____

PLEASE READ DIRECTIONS BEFORE BEGINNING: PLEASE ASSIGN TO STUDENTS YOU NOW TEACH IN ONE OR MORE BASIC SKILLS, ONE OF THE FOLLOWING RATINGS:

"1" This student is below minimum standards and definitely needs remedial help.

"2" This student is close to minimum standards and may or may not need remedial help.

"3" This student is above minimum standards and definitely does not need remedial help.

NAME	BASIC SKILL		
	READING	LANGUAGE ARTS (WRITING)	MATH

Appendix H

Data Entry Sheet - MEC Improving Basic Skills Project

School District ______________________

Grade Level ____ ____/____ Mo. Yr.

Part A | Part B

ast ame	First Name		Spec Grp.	Sex	Min Stat *	Lang Stat **	Grades***			Test Scores			Teacher Ratings[t]		
							Read	Lang Arts	Math	Read	Lang Arts	Math	Read	Lang Arts	Math

* If minority status - enter 1
If non-minority status - enter Ø

** If non-English speaking - enter 1

*** A = 4
B = 3
C = 2
D = 1
F = Ø

t Student performance considered below standard --- enter 1

Student performance considered about standard --- enter 2

Student performance considered above standard --- enter 3

✓ to comply with confidentiality procedures detach and forward only Part B to MEC/ETS.

Data Entry Sheet - MBC Improving Basic Skills Project: Directions

Part A calls for student names for any one grade level to be listed in whatever order is convenient for local purposes. To comply with confidentiality procedures, after the entire form (Parts A & B) is completed, Part A is to be detached from Part B by cutting along the dotted line. Only Part B is to be forwarded outside of the school district.

Part B calls for several pieces of information on each student, each of which is described here.

"Spec. Group" - identifies the student as belonging to one of the following groups:

Title I students (not on CAI) - enter "1"
Title I students (on CAI) - enter "2"
Special education students declared eligible for basic skill testing - enter "3"
(Those special education students who were not explicitly declared eligible for basic skill testing should NOT be named on the list)

"Sex" - identifies the student as: female - enter F or male - enter M

"Min Stat" identifies the student's minority group status as
belonging to a minority group - enter 1
not belonging to a minority group - enter Ø

"Lang Stat" identifies the language status of the student as
from a home in which English is NOT spoken - enter 1
or
from a home in which English is spoken - enter Ø

"Grades" call for the most recently available semester grades in reading, language arts or writing, and mathematics, which should not include Algebra. Grades should be entered as numbers according to the following scale:

A = 4	D = 1	Pluses or minuses should be overlooked for purposes of this project.
B = 3	F = 0	
C = 2		

"Test scores" call for publishers' scaled scores from all basic skill objective tests and total scores on essays. "Read" refers to total reading scores, "Lang Arts" refers to the "Language Arts" score on the standardized tests and the total score on the writing exercise. "Math" refers to total math scores.

"Teacher Ratings" are to be transcribed from the class lists completed by the teachers, on which each student is rated as:

"1" below standard
"2" at or close to standard
"3" above standard

If two or more teachers happen to submit the same rating on a student in a given skill area, use the rating. If they submit different ratings, enter a "2".

The completed forms should be returned to your local district facilitator for the project.

Appendix I

REGULATIONS FOR THE IMPLEMENTATION OF THE BASIC SKILLS IMPROVEMENT POLICY

Commonwealth of Massachusetts
Department of Education
January 23, 1979

Publication #11182 approved by Alfred C. Holland, State Purchasing Agent.

603 CMR: DEPARTMENT OF EDUCATION

603 CMR 40.00: REGULATIONS FOR THE IMPLEMENTATION OF THE BASIC SKILLS IMPROVEMENT POLICY

Regulations governing the implementation of the Basic Skills Improvement Policy were adopted by the Massachusetts Board of Education on January 23, 1979.

Introduction

The purpose of these regulations is to implement the "Policy on Basic Skills Improvement" adopted by the Board of Education on August 29, 1978 and to assist students in achieving mastery of basic skills prior to high school graduation through the provision of appropriate curriculum, instruction, and evaluation. These regulations shall apply to all public school districts providing educational services to students in all or part of grades kindergarten through twelve in the Commonwealth of Massachusetts. These regulations are promulgated by the Board of Education pursuant to its powers under Section 1G of Chapter 15 of the General Laws which directs that "the board shall establish minimum educational standards for all courses which public schools require their students to take" and which authorizes the board to "collect and maintain information from any public school system in the Commonwealth relevant to its work..." In implementing these regulations, public school districts shall place emphasis on diagnosing learning needs and adjusting the regular curriculum to meet these needs.

40.01: Definitions

The various terms as used in these regulations are defined below:

(1) Advisory Committee shall mean the Advisory Committee on Basic Skills Improvement established by the Board of Education pursuant to Section 40.11 of these regulations.

(2) Annual Report shall mean the Annual Report on Basic Skills Improvement Programs which public school districts shall submit to the Department of Education pursuant to Section 40.09 of these regulations.

(3) Basic skills shall mean the following skills: reading, writing, listening, speaking and mathematics.

(4) Basic skills improvement programs shall mean the programs which public school districts are required to establish by Section 40.02 of these regulations.

(5) Early elementary level shall mean all or part of grades kindergarten through three.

(6) Early elementary and later elementary level program plans shall mean the written documents public school districts shall prepare pursuant to Section 40.06 of these regulations.

(7) Evaluation instruments shall mean the tests or assessment techniques that public school districts shall use to determine whether students have or have not achieved minimum standards.

(8) Later elementary level shall mean all or part of grades four through six.

(9) Minimum standards shall mean (a) the objectives in the basic skills and (b) the level of achievement for the objectives in the basic skills that public school districts shall establish for students for the early elementary, later elementary, and secondary levels pursuant to these regulations.

(10) Parent shall mean a father, mother, or legal guardian.

(11) Public school district shall mean city, town and regional school committees; local and district trustees for vocational education and boards of trustees of county agricultural schools created pursuant to General Laws, Chapter 74; Superintendents of Smith's Agricultural School established by Chapter 151 of the Special Acts of 1918; and the governing board of any other public school providing educational services to students in all or part of grades kindergarten through twelve in the Commonwealth of Massachusetts. The public school district is responsible for the overall implementation of these regulations, but it may delegate to its public school administration the specific tasks required by the regulations, as appropriate.

(12) Regional Education Centers shall mean the six operational field offices of the Department of Education located in specified geographical regions of the Commonwealth.

(13) Regional Education Councils shall mean the councils established by the Board of Education pursuant to the "Revised Board of Education Policy on Regional Education Councils."

(14) Secondary level shall mean all or part of grades seven through twelve.

(15) Secondary level program plans shall mean the written documents public school districts shall prepare pursuant to Section 40.07 of these regulations.

40.02: Basic Skills Improvement Program

(1) Each public school district shall establish a basic skills improvement program which shall include the establishment of minimum standards and provision for evaluating student achievement of minimum standards required by Sections 40.03 to 40.05 of these regulations, and the establishment and implementation of the early elementary, later elementary and secondary level program plans required by Sections 40.06 and 40.07 of these regulations.

(2) Each public school district shall provide for genuine public participation in the development and periodic review of its basic skills improvement program. At all grade levels, public participation shall include: employers; parents, including Title I parents, of students enrolled in the public school district; teachers and administrators employed by the public school district; and the general public. In addition, at the secondary level, public participation shall include secondary school students enrolled in the public school district.

40.03 Establishment of Minimum Standards

(1) Each public school district shall establish minimum standards in each of the basic skills for the following grade levels, if applicable: early elementary, later elementary and secondary. At the secondary level the minimum standards established by each public school district shall include, but not necessarily be limited to, the basic skills objectives set forth in Section 40.04 of these regulations.

(a) By August 1, 1980, each public school district shall establish minimum standards in the basic skills of reading, writing and mathematics.

(b) By August 1, 1981, each public school district shall establish minimum standards in the basic skills of listening and speaking.

(2) Each public school district shall determine the specific grade or grades within the early elementary, later elementary and secondary levels for which minimum standards shall be established.

40.04. Secondary Level Minimum Standards

The secondary level minimum standards shall include, but not necessarily be limited to, the following basic skills objectives:

(1) Reading.
- (a) Basic Word Meaning
  1. Identify the meaning of commonly used words within a sentence that does not provide clues to the meaning of the word
  2. Identify the meaning of a word within a sentence that provides clues to the meaning of the word
- (b) Literal Comprehension
  1. Identify the meaning of a written phrase, clause, sentence, or paragraph
  2. Demonstrate the ability to follow directions
  3. Identify the main idea, supporting details, and conclusion of a paragraph
  4. Recognize the sequence of events or ideas in a written passage
  5. Identify information on a chart, map, or graph
- (c) Interpretive Comprehension
  1. Draw conclusions implied in a paragraph or passage
  2. Identify cause and effect relationships implied in a paragraph or passage
  3. Predict an outcome implied in a paragraph or passage
- (d) Evaluative Comprehension
  1. Identify a statement as fact or opinion
  2. Identify the writer's purpose in a paragraph or passage written to inform or persuade
- (e) Locating Information
  1. Use the parts of a book
  2. Locate information in a variety of sources

(2) Writing. Given the opportunity to use a dictionary, students, through their own writing samples, will demonstrate:
- (a) Knowledge of the subject
  1. The writer has something to say
  2. Ideas are suported with relevant details
- (b) Clear and consistent purpose
- (c) Organization
  1. Ideas are related
  2. Ideas progress logically from one point to another
- (d) An awareness of the intended reader

(e) Precise word choices
1. Words appeal to the reader's senses
2. Words suit the purpose
3. Words are appropriate for the intended reader
(f) Fulfillment of the purpose
1. Adequate information is provided
2. The writing is free of irrelevancy
3. The conclusion reemphasizes the purpose
(g) Correct capitalization and punctuation
(h) Correct spelling
(i) Legible handwriting
(j) Complete sentences
(k) Standard use of nouns, pronouns, verbs, adjectives, and adverbs
(l) Agreement of subject and verb

(3) Listening.
(a) Basic Listening Skills
1. Recognize words and phrases used by the speaker
2. Indicate why the speaker can or cannot be understood
(b) Understanding What You Hear
1. Understand spoken words and ideas
2. Identify and understand main ideas
3. Associate important details with main ideas
4. Understand descriptions of events and experiences
5. Understand speaker's purpose
(c) Using What You Hear
1. Understand and respond to survival words used in emergency situations
2. Summarize information and draw conclusions
3. Recognize when words and phrases are used to convince or persuade
4. Follow straightforward directions

(4) Speaking.
(a) Basic Oral Communication Skills
1. Use words and phrases appropriate to the situation
2. Speak loudly enough to be heard by a listener or group of listeners
3. Speak at a rate the listener can understand.
4. Say words distinctly
(b) Planning, Developing and Stating Spoken Messages
1. Use words in an order that clearly expresses the thought.
2. Organize main ideas for presentation
3. State main ideas clearly
4. Support main ideas with important details
5. Demonstrate knowledge of standard English usage
(c) Common Uses of Spoken Messages
1. Use survival words to cope with emergency situations
2. Speak so listener understands purpose
3. Ask for and give straightforward information
4. Describe objects, events and experiences
5. Question others' viewpoints

(5) Mathematics

(a) Number and Numeration Concepts

1. Recognize number symbols (17, eighteen), whole numbers (34), fractions ($\frac{1}{2}$), decimals (3.75), and powers of 10 (10^2)
2. Identify odd and even numbers
3. Put numbers in numerical order
4. Recognize equivalent fractions $\left[\frac{2}{4} = \frac{1}{2}\right]$

(b) Arithmetic Computation

1. Add, subtract, multiply, and divide whole numbers (4069 + 81 + 123, 254 x 17, 16,300 ÷ 100)
2. Add and subtract mixed numbers $\left[\begin{array}{r} 2\frac{1}{2} \\ \hline 1\frac{1}{4} \end{array}\right]$
3. Multiply whole numbers or money by fractions (halves, quarters, thirds)
4. Add, subtract, multiply, and divide decimal numbers like money
5. Change a fraction to a decimal ($\frac{1}{4}$ to .25)
6. Find a percent of a number in situations such as simple interest, discounts, commissions, and taxes
7. Use ratio and proportion (mixtures, recipes, scale drawings)
8. Use simple formulas ($A = \ell \times w$)

(c) Estimation and Approximation

1. Round off numbers to a specified place
2. Approximate the answer to a computation problem (including discounts and percentages)
3. Estimate length, weight/mass, capacity, time, temperature, area, and volume
4. Estimate with money

(d) Measurement and Geometry

1. Choose an appropriate unit of measurement in the U.S. customary system (for example, feet, pounds, and gallons)
2. Choose an appropriate unit of measurement in the metric system (for example, meters, kilograms, and liters)
3. Choose an appropriate measurement instrument involving both U.S. customary and metric units
4. Convert common measurements within the same system
5. Read a scale drawing
6. Use a map to compute highway distances
7. Relate total cost and cost per unit
8. Compute by using temperature
9. Compute by using time
10. Identify right angles and parallel, perpendicular, and intersecting lines like those in a street map
11. Recognize that an object has the shape of a square, rectangle, triangle, or parallelogram
12. Identify the radius, diameter, and center of a circle
13. Recognize that an object has the shape of a cube, cylinder, or sphere
14. Find the perimeter of a triangle, square, and rectangle
15. Find the area of a triangle, square, and rectangle
16. Find the volume of a cube or other rectangular solid

(e) Graphs and Tables
1. Read a table
2. Interpret a bar graph
3. Interpret a circle graph
4. Interpret a line graph

(f) Prediction of Events and Statistics
1. Understand probabilities like those used in weather forecasting or lotteries. (the chance something will or will not happen)
2. Find and use averages (mean and median) for a group of numbers

40.09: Evaluation of Basic Skills Achievement

(1) Each public school district shall evaluate each student's achievement of minimum standards at least once during the early elementary, later elementary and secondary levels.

(a) Evaluation of student achievement of minimum standards in reading, writing and mathematics shall begin no later than the end of the 1980-81 school year.

(b) Evaluation of student achievement of minimum standards in listening and speaking shall begin no later than the end of the 1981-82 school year.

(2) Each public school district shall determine at which grade or grades at the early elementary, later elementary, and secondary levels students shall be evaluated for the achievement of minimum standards, provided that secondary level students shall first be evaluated no later than the beginning of grade nine.

(3) Once a public school district has determined at which secondary grade level to begin the evaluation of student achievement of minimum standards, it shall provide for an evaluation of student achievement of minimum standards at least once at each succeeding grade level until the student has achieved the minimum standards. A secondary level student who has achieved the minimum standards for a basic skill need not be evaluated again in that skill.

(4) At the early and later elementary levels, each public school district shall determine which evaluation instruments shall be used to evaluate student achievement of minimum standards.

(5) At the secondary level, each public school district shall have the option of using one or more of the following evaluation instruments to evaluate student achievement of minimum standards:

(a) Evaluation instruments available from the Department of Education;
(b) Commercially available evaluation instruments approved by the Department of Education; or
(c) Locally utilized or developed evaluation instruments approved by the Department of Education as being comparable to either (a) or (b).

(6) In accordance with Section 4.03 of the Chapter 622 Regulations Pertaining to Access to Equal Educational Opportunity as adopted by the Board of Education, a public school district shall not use an evaluation instrument which discriminates on the basis of race, color, sex, religion or national origin.

(7) After an evaluation of a student's achievement of minimum standards, each public school district shall give the student and his or her parent the opportunity to review the evaluation instruments used and the complete record of the student's achievement level. The complete record of a student's achievement level shall include, but not necessarily be limited to, the corrected responses of the student on the evaluation instruments used, and shall be maintained as part of the student's temporary record pursuant to the Student Records Regulations as adopted by the Board of Education.

10.06: Early Elementary and Later Elementary Level Program Plans.

(1) The early elementary and later elementary program plans shall include, but not necessarily be limited to, the following components:

(a) Provision for participation of parents, teachers, administrators, employers, and the general public in the establishment and periodic review of minimum standards and the overall early elementary and later elementary program plans;

(b) Establishment of minimum standards for each basic skill;

(c) Designation of appropriate evaluation instruments;

(d) Provision for appropriate follow-up instructional programs and services to students who have not achieved minimum standards;

(e) Plans for monitoring, evaluating and modifying, as appropriate, the early elementary and later elementary program plans that have been established. These plans shall include such activities as: reviewing elementary programs and curriculum; making necessary program and curriculum changes; conducting appropriate staff development efforts, planned in cooperation with affected staff; and implementing appropriate plans for monitoring and evaluating the outcome of these efforts.

(2) Each public school district shall affirm to the appropriate Regional Education Center, on a form and in a manner prescribed by the Department of Education, that it has established early elementary and later elementary program plans and shall make these plans available to the public.

(a) By August 1, 1980, each public school district shall affirm that it has established early elementary and later elementary program plans in reading, writing and mathematics.

(b) By August 1, 1981, each public school district shall affirm that it has established early elementary and later elementary program plans in listening and speaking.

(3) Each public school district shall implement its early elementary and later elementary program plans for reading, writing and mathematics no later than October 1, 1980 and for listening and speaking no later than October 1, 1981.

40.07: Secondary Level Program Plan

(1) The secondary level program plan shall include, but not necessarily be limited to, the following components:

(a) A description of how parents, employers, teachers, administrators, students and the general public participated in the establishment of the public school district's minimum standards and the overall secondary level program plan;

(b) The secondary level minimum standards established by the public school district and how these secondary level minimum standards include the secondary level basic skills objectives set forth in Section 40.04 of these regulations;

(c) A listing of all evaluation instruments designated by the public school district. If an evaluation instrument designated is not one of those made available by the Department of Education or on a Department of Education approved list, the public school district shall submit the following:

1. a copy of the selected evaluation instruments, with supporting technical information;
2. a detailed analysis showing the content match between the instruments and the minimum standards set forth in Section 40.04 of these regulations.

(d) The grade level or levels that the public school district has established for the initial evaluation of secondary level student achievement of minimum standards;

(e) A description of how the public school district will provide, and make generally available to the public, information about its secondary level program plan, including such information as: where in the curriculum basic skills are covered and what diagnostic and follow-up instructional services are available;

(f) A description of the follow-up instructional programs and services that the public school district will provide to students not achieving minimum standards;

(g) Assurances that:

1. present school programs and curricula will be reviewed for all grades in light of the school district's minimum standards and appropriate modifications will be made as necessary;
2. curriculum and instructional services will be reviewed following analysis and reporting of student achievement data;
3. staff development efforts will be planned in consultation with staff affected and will relate to implementation of basic skills improvement programs.

(2) Each public school district shall submit for approval to the appropriate Regional Education Center, on a form and in a manner prescribed by the Department of Education, its secondary level program plan and shall make this plan available to the public.

(a) Each public school district shall submit its secondary level program plan for reading, writing and mathematics no earlier than February 1, 1980 and no later than August 1, 1980.

(b) Each public school district shall submit its secondary level program plan for listening and speaking no earlier than February 1, 1981 and no later than August 1, 1981.

40.08: Approval of Secondary Level Program Plan

(1) Each Regional Education Council or a committee designated by the Council shall review the secondary level program plan of each public school district within its region. Within sixty (60) calendar days of receipt of a public school district's plan, the Regional Education Council shall recommend to the Commissioner of Education approval or disapproval of the secondary level program plan. This recommendation will be based on whether or not the plan contains the information required by Section 40.07 (1) and otherwise complies with these regulations.

(2) Within thirty (30) calendar days of receipt of the Regional Education Council's recommendation on a public school district's secondary level program plan, the Commissioner of Education shall approve or disapprove the secondary level program plan and notify the public school district of its action. If a secondary level program plan is not approved, this notification shall state the reasons for the disapproval.

(3) Each public school district shall implement the secondary level program plan for reading, writing and mathematics approved by the Commissioner of Education within thirty (30) days of notification of such approval or by October 1, 1980, whichever occurs later.

(4) Each public school district shall implement the secondary level program plan for listening and speaking approved by the Commissioner of Education within thirty (30) days of notification of such approval or by October 1, 1981, whichever occurs later.

(5) Upon notification that a secondary level program plan has been disapproved, a public school district shall have thirty (30) calendar days to submit an amended secondary level program plan to the appropriate Regional Education Center. Within thirty (30) calendar days of receipt of an amended secondary level program plan, the Regional Education Council or a committee designated by the Council shall review the amended secondary level program plan and recommend to the Commissioner of Education that the amended plan be either approved or disapproved. If the amended secondary level program plan is approved, the public school district shall implement this plan within thirty (30) calendar days of notification.

(6) Each public school district may change its approved secondary level program plan by submitting an amendment to its program plan to the appropriate Regional Education Center. The Regional Education Council

or a committee designated by the Council shall review the amendment to the program plan and recommend to the Commissioner of Education that the amendment be approved or disapproved in accordance with the provisions of this section. No new secondary level program plan shall be implemented in place of an approved secondary level program plan until it has been approved by the Commissioner of Education.

40.09: Annual Report on Basic Skills Improvement Programs

(1) Beginning with the 1980-81 school year, each public school district shall submit an Annual Report, on a form and in a manner prescribed by the Department of Education, on the implementation of its basic skills improvement program. The Annual Report shall include the number and percentage of students by race, sex, and linguistic minority who have and have not achieved the minimum standards for each basic skill established by the public school district for the early elementary, later elementary and secondary levels. The Annual Report shall be submitted to the appropriate Regional Education Center by August 31 for the school year ending the previous June 30.

(a) For the 1980-81 school year, and each year thereafter, each public school district shall decide which grade or grades within the early and later elementary levels shall be used to report the number and percentage of students who have and have not achieved minimum standards.

(b) For the 1980-81 school year and each year thereafter, each public school district shall decide which grade or grades within the secondary level shall be used to report the number and percentage of students who have and have not achieved minimum standards, provided that, for the 1982-83 school year and each year thereafter, each public school district shall report the number and percentage of students who have completed eleventh or twelfth grade and who have and have not achieved minimum standards.

(2) Each public school district shall release and make generally available to the public the Annual Report required by subsection (1), and shall take reasonable steps to insure that the people residing in the public school district are informed of the content and availability of the Annual Report.

40.10: Transfer Students, Students in Special Education Programs, and Students of Limited English Ability

(1) A public school district may exempt any student who transfers into its school district after having completed the tenth grade in another district from the annual reporting requirements of Section 40.09.

(2) The evaluation team responsible for developing a student's Individual Educational Plan in accordance with General Laws, Chapter 71B and the Chapter 766 Regulations shall determine whether a student in need of special education shall be evaluated for achievement of the minimum standards provided for by these regulations, in addition to the

objectives contained in the Individual Educational Plan, and whether the result of such an evaluation shall be included in the public school district's annual report submitted pursuant to Section 40.09. The evaluation team's determination shall be included in the student's Individual Educational Plan.

(3) Each public school district shall exempt, from all provisions of these regulations, students who have limited English ability as defined by General Laws, Chapter 71A. Notwithstanding the previous sentence, students who have completed an appropriate transitional bilingual education program in accordance with General Laws, Chapter 71A and the Regulations for Use in Administering Programs in Transitional Bilingual Education adopted by the Board of Education shall be considered as any other students for the purpose of these regulations.

40.11: Advisory Committee on Basic Skills Improvement

(1) The Board of Education shall establish an Advisory Committee on Basic Skills Improvement to advise on matters related to the implementation of the Board of Education's Policy on Basic Skills Improvement and these regulations.

(a) Until June 30, 1979, the Policy Committee on High School Graduation Requirements established by the Board of Education on June 28, 1977 shall serve as the Advisory Committee.

(b) On or before July 1, 1979, the Board of Education shall appoint no less than 25 people to serve as voting members of the Advisory Committee. Members shall be appointed for two year terms, and may be reappointed for an additional two year term. No member may serve for more than two terms. If an Advisory Committee member does not complete the appointed term, the Board of Education may appoint a member to fill the unexpired term and this member may, upon reappointment, serve two consecutive full terms. Subject to appropriation, members of the Advisory Committee shall be reimbursed for expenses necessarily incurred in the performance of their official duties.

(2) Beginning in 1980, the Advisory Committee shall submit an annual report to the Board of Education on or before September 30 of each year to cover the preceding school year.

40.12: Waiver

The Board of Education may, upon written application of a public school district and recommendation of the appropriate Regional Education Council and Commissioner of Education, grant a waiver of part or all of the requirements of these regulations for good cause.

40.13: Severance Clause

If any provision contained in these regulations or the application thereof to any person or circumstance is held invalid by a court of competent jurisdiction, the remaining provisions of these regulations or the application of the provision in question to other persons not similarly situated or to other circumstances shall not be affected thereby:

Some Points to Consider
to
Eliminate Sex and Ethnic Bias in Tests

I. Some "Dos and Donts" with Respect to Sex-Fairness

A. No item in any test should include sexist language or description. Every attempt should be made to avoid

1. the use of demeaning stereotypes and modifiers
2. the use of male oriented generic terms wherever possible without making wording complex or stilted
   a. avoid generic "he" by use of plurals or change in structure
   b. for an anonymous person, sometimes use he, sometimes she
   c. avoid male oriented terms for occupations and materials (e.g., police officer, not policeman; synthetic, not manmade)
3. labelling any attribute, trait, interest, desire or liking as masculine or feminine, rather than just human
4. describing women or girls by their physical or personality qualities when men or boys are described by their accomplishments

B. In each test, every attempt should be made to include items that

1. mention the contributions of women to civilization
2. depict women and girls in leadership, professional and managerial roles and as active participants, not just as receivers or observers
3. treat women with respect and their issues and ideas with importance

C. In each test, every attempt should be made to achieve balance in the way males and females are presented.

1. If one item presents a female in a stereotypic role, it should be balanced by another presenting her in a more nontraditional role or a male in a traditionally female role
2. The ways in which males and females are referred to should be parallel
3. There should be reasonable balance in the number of items about males and females

4. There is balance in the ways in which males and females are presented with respect to occupations, accomplishments, roles, and positive or negative traits

5. In a history test, for example, where male referrants will necessarily be predominant, some items referring to women and their contributions should be included

II. Some Dos and Donts with Respect to Ethnic Minorities

A. In writing test items, avoid

1. the description of members of ethnic minorities in terms of physical, personality or interest stereotypes

2. the use of language which might be considered derogatory by minority groups

3. the use of words which might have a different meaning in different cultural settings or dialects

4. the use of subject matter likely to be unfamiliar to minority groups while familiar to the majority

5. the use of esoteric vocabulary or complex sentence structure when that is not what is being tested

6. the use of material presenting highly controversial or prejudiced points of view

B. In each test, every effort should be made to include

1. reading material relevant to the interests and concerns of ethnic minorities or stressing the positive values of cultural diversity

2. reading material presenting positive role models from minority groups or which discusses minority contributions to science, history, government and the arts

3. a balanced use of minority and majority related subject matter and names of well-known personalities in the context of items testing such things as grammatical usage

4. the occasional use of names which suggest membership in different ethnic groups for the person referred to in the context of a problem or question (e.g., Carlos or Juanita instead of John or Mary in a math problem; Mrs. Martinelli or Mr. Krolikowski as the school principal instead of Miss Smith or Mr. Jones)