

DOCUMENT RESUME

ED 148 833

TR 006 507

AUTHOR Fremer, John; Dwyer, Carol A.
TITLE Setting Standards for Basic Skills Reading Assessment.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE 6 May 77
NOTE 18p.; Paper presented at the Annual Meeting of the International Reading Association (Miami, Florida, May, 1977)

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Basic Skills; *Criterion Referenced Tests; Diagnostic Tests; Educational Objectives; Elementary Secondary Education; *Evaluation Methods; Measurement Goals; *Reading Ability; Reading Diagnosis; *Reading Tests; School Responsibility; Speech Skills; *Standards; Student Evaluation; Test Construction; *Testing Programs

ABSTRACT

State and school district testing activities involving basic reading competency requirements are discussed. The purposes of minimum competency testing in reading include the identification of students who need remedial assistance, and the evaluation of secondary school students regarding graduation requirements. Important considerations regarding test content include: its relationship to the assessment purposes, the role of task analysis, the priorities of particular skills, the type of test, the ways to include teacher judgment, the practical vs. academic flavor of the test, the methods for presenting very difficult materials, and the need for a more developed theory of reading. Several general properties of good systems for setting standards are presented: (1) The school's responsibility toward the needs of the students and of the society must be recognized. (2) Management standards regarding planning, delegation of responsibility, and review or quality control should be maintained. (3) Expectations must be realistic. (4) Standards should be administered with adequate student preparation and feedback. (5) Standards should be applied consistently and should be acceptable to other institutions, so that credits will be transferred. (6) Standards should be administered fairly so that the involved students feel they have been well served. (Author/MV)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. Nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Fremer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM.

Setting Standards for Basic Skills Reading Assessment¹

John Fremer and Carol A. Dwyer
Educational Testing Service, Princeton, New Jersey

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

BASIC SKILLS ASSESSMENT AROUND THE NATION

- This handout summarizes some state and local district activity involving competency requirements in the Basic Skills area. Since Fred Finch has done an excellent job reviewing this activity I won't spend time discussing the report.
- I do want to make two related comments on the trends shown in the handout. I believe these comments are consistent with yesterday's presentations:
 1. Many of the new competency requirements have been adopted without adequate planning and without careful consideration of consequences.
 2. Some of the legislation and some of the state and local board rulings are clearly unsound technically and probably unworkable educationally. Many of these flaws could easily be worked out in a pilot study.
- Despite my concerns, though, about some of the ways the movement is developing, I still feel it has great potential. Moreover, I believe that we can make a great improvement over current practice merely by pulling together some of the best of current practice.

¹ Paper presented at the annual meeting of the International Reading Association, Miami, May 1977. [Part of Perspectives in Reading Conference--"Legislation and Reading Competency--Understanding the Nationwide Trend Toward Performance Assessment Before and For Graduation."]

ED148833

006507

TM

FOUR PARTS TO MY PRESENTATION

- Purpose of Minimum Competency Measurement in Reading
- Content of Minimal Competency Reading Tests
- General Properties of Good Standards Setting Systems
- Methods of Setting Standards

Purpose of Minimum Competency Measurement in Reading

One general purpose is to identify students who need remedial assistance to reach a level of reading skill that will facilitate their future development. Discussion of the future goal for which a minimum reading level should be set tends to focus on concepts such as

- ability to participate in additional education (either still at school or post high school level)
- ability to function in society as a consumer, producer, citizen, etc.

Important to keep in mind the fact that acquiring basic skills clearly doesn't guarantee adequate functioning.

- Many who have highly developed skills are not doing well.
- Many have been very successful without skills.

This purpose raises issues such as:

- Can there be a single minimum standard?

These two purposes raise many questions including:

- What about students who can't meet the minimum?
- What about students beyond the minimum?

Another clear purpose, very evident in state legislation, is that of limiting access to the high school diploma to those with at least some minimal level

of reading skill. This raises many additional issues including:

- How many "remedial" students can a school afford?

—not just financial but political issue

(Example of unfortunate political situation: Local CRT project that went bad.)

We were working with a local school district to develop a CRT system for instructional management. The intention was to provide each teacher at the early elementary grades with objectives and associated test questions.

We had to consider

- Which objectives should be included -- at which grades?
- Unfortunately, teaching staff reached the conclusion that the Superintendent planned to change emphasis to a teacher accountability system. Don't know if this was true, but that is what teachers thought.
- So, as a defense, teachers set objectives and standards that they were sure they could meet.

--Entering first grade students were tested to see how they stood on end of first grade objectives

--85% had already mastered the first grade objectives

This kind of active guerilla warfare can easily occur in any setting when an assessment system is imposed on the schools.

CONTENT OF MINIMAL COMPETENCY READING TESTS

This presentation is addressed primarily to setting standards but some general points can be made regarding test content.

1. Relationship to Purpose - Content should be consistent with purpose for minimal competency testing. If primary interest is in seeing whether students can read materials they will encounter outside of school, test should cover such material. At the secondary level, test should be reading comprehension. Then if students cannot meet the school's standard, follow-up evaluation should be carried out using tests as only one tool and building on past knowledge about student.
2. Role for Task Analysis - Analyses of reading demands of school and adult life need to be carried out and studied. Actually we have more to go on in the area of reading than in mathematics, writing, or other areas.

As part of USOE's Targeted Research and Development Program in Reading, an extensive national survey of adult reading activities was carried out in the years 1970-75. This included door-to-door surveys to find out what adults were reading and for what reason.

- Exercises developed and administered door-to-door to national sample

This study carried out by ETS can be supplemented by APL, NAEP, and a number of state and local district projects.

3. Decision Re Skills Balance - What should be the place of decoding skills, literal comprehension, and higher level inference and evaluation?

Generally literal comprehension:

- Our BSA - 50% literal comprehension
- 40% straightforward inference
- 10% evaluation

In considering skills balance, aim for an "easy" test.

- A. We tend to make tests too difficult for best measurement.
- B. It will be very difficult to explain a low standard.
- C. In most places our interest is in lower part of class.

4. Form of Test - Should objectively scannable tests be used?
In combination with other assessment and evaluation techniques?

5. Link to Teacher Judgment - Need to think about ways that test based information can be combined with teacher judgments.

(Applies to 4 and 5)

Need overall decision - What will be the basic approach? Then you need alternative approaches when the basic approach is not feasible.

6. Flavor of Test - Practical, real-world, applied materials only?
Balance of real world and academic?

7. What About Unreadable Materials? - Does Form 1040 belong on a minimal competency reading test?

(Applies to 6 and 7)

- Typical survey reading test based on school material
- We have some tests now that have "real life" content--
APL, LA City, Everyday Skills, BSA
- Probably will reach some kind of balance in tests
- Those unreadable materials pose an interesting problem--
John Geyer mentioned that many materials that are essential go off the readability scales. This shows up in indirect testing.
- I endorse John Geyer's appeal re research and rewriting.
It could also help if we could all try to publicize the problems these materials cause.

8. Need for More Developed Theory of Reading? - Does atheoretical equal amoral?

A question I carried away from the AERA meeting on topic --
What should minimum competency reading tests measure?

I have assumed that the tests should contain materials and questions that were judged appropriate by reading people and community people. The sample of content should be based on evidence regarding what adults read. The kinds of questions related to our knowledge or assumptions, if necessary, about what adults are trying to learn from materials. Questions must be technically sound.

I have heard a strong call for much more rigorous theoretical analyses and for highly systematic test question development rules.

Research -- Let's have it

Mechanistic item development -- I react to
that the way I do to paint-by-numbers
as an art form.

DEFINITIONS — GOING BACK TO THE BIBLE¹

- Want to touch on the issue of definitions and to give you some background on the development of "standards," "passing scores," and "cutting scores."
- As my reference book, I am using The Bible (King James Version), Old Testament, Book of Judges, Chapter 12. (In case some of you don't remember that chapter, it deals with the Gileadites and Ephraimites.)
 - Gileadites were holding the passes (passages) of Jordan, Ephraimites wanted to go through these passes. [They wanted to earn a passing score.]
 - Gileadites had developed a one item test that they had a lot of confidence in.
 - If they thought someone was an Ephraimite
 - "Then said they unto him, say now Shib - bo - leth: and he said Sib - bo - leth for he could not frame to pronounce it right."
 - Those who said Sib-bo-leth and thus did not achieve a (Jordan) passing score were slain
 - "... and there fell at that time forty and two thousand."
 - My ETS colleague, Bill Angoff, says that the Ephraimites plus some Gileadites with speech handicaps and some other non-narrative speakers probably had their heads cut off. He feels that this was the origin of the notion of a "cutting score" on a test.

General Properties of Good Standards Setting Systems

1. Ethics/Responsibility. Competency standards for the diploma should be developed by a process that gives major attention to the needs of students but which recognizes an institution's responsibility to societal needs.

--There's an appeal for high school diplomas or some related credential to serve as evidence of learning. Yet we want our schools to open doors not close them.

--I thought it worth stating as an issue of ethics.

2. Management Standards. Competency standards should be developed using good management standards or procedures, e.g., with adequate attention to:

planning,
delegation of responsibility, and
review and quality control.

Management standards apply to the steps taken in the design and implementation of competency standards for the diploma, so that necessary decisions are made by qualified people in a manner that provides for consistent quality.

An important first step in evaluating the adequacy of an institution's approach to standard-setting procedure is to determine whether overall responsibility for this task has been defined.

The particular individual(s) who hold^(s) this responsibility can be expected to vary depending on the nature of the program, department, division, or other part of the institution within which evaluations take place. Although someone with training and experience in the field of evaluation would be best for the job, those without such background are often selected. In such instances, the use of trained evaluation consultants will be essential.

Even when overall responsibility for establishing performance standards can be placed in the hands of a skilled evaluator, many

other individuals should contribute to the standard-setting process. For this purpose, it will be useful to seek assistance not only from faculty and students but also from members of the larger community.

The job of designing and developing competency standards cannot be handled adequately if decisions about particular students must be made under great time pressure. Providing adequate calendar time, and also adequate staff for the development of competency standards, will make it possible to plan and carry out pilot projects to try out standards setting systems. Procedures can be established for monitoring the application of standards and for following up on the students afterwards. Most institutions will find it useful to adopt an iterative process for developing competency standards. The best available sources of information should be used when setting standards for the first group of students tested. The experience of conducting the initial evaluations, though, should become part of the background for future evaluations.

3. Realism. Standards must be set at realistic levels.

Critics of U.S. high schools often criticize the lack of rigor associated with the evaluation of students' learning experiences. Little attention has been paid to the reverse problem-- setting unrealistically high standards. This problem is most likely to occur when the individuals participating in the setting or application of standards do not have sufficient direct knowledge of the performance of typical students. It is the responsibility

of the institution to judge what is reasonable to expect of students before making decisions regarding competency standards.

- We had an example yesterday—as I heard Harry Handler, the LA City School Board thought all questions should be above average in difficulty because they wanted students to be above average. [Setting high standards doesn't insure meeting them.]
- Clearly we do need high standards. We don't want minimums to become maximums. The direction to move in, though, has to be that of developing standards at more than one level. Many possibilities:

<u>2 level</u>	<u>3 level</u>	multilevel scales
honors	honors	
minimum	average	
	minimum	

A partial sample of a multilevel scale is the language proficiency scales developed by the U.S. Foreign Service Institute and used for many years with Peace Corps volunteers:

Elementary Proficiency

Speaking-1. Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language experience can understand simple questions and statements, allowing for slowed speech; repetition or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at the S-1 level should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

Reading-1. Able to read some personal and place names, street signs, office and shop designations, numbers, and isolated words and phrases. Can recognize all the letters in the printed version of an alphabetic system and high-frequency elements of a syllabary or a character system.

Limited Working Proficiency

Speaking-2. Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations including introductions and casual conversations about current events, as well as work, family, and autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on non-technical subjects (i.e., topics which require no specialized knowledge) and has a speaking vocabulary sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.

Reading-2. Able to read simple prose, in a form equivalent to typescript or printing, on subjects within a familiar context. With extensive use of a dictionary can get the general sense of routine business letters, international news items, or articles in technical fields within his competence.

Minimum Professional Proficiency

Speaking-3. Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabulary is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.

Reading-3. Able to read standard newspaper items addressed to the general reader, routine correspondence, reports and technical material in his special field. Can grasp the essentials of articles of the above types without using a dictionary; for accurate understanding moderately frequent use of a dictionary is required. Has occasional difficulty with unusually complex structures and low-frequency idioms.

Full Professional Proficiency

Speaking-4. Able to use the language fluently and accurately on all levels normally pertinent to professional needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

Reading-4. Able to read all styles and forms of the language pertinent to professional needs. With occasional use of a dictionary can read moderately difficult prose readily in any area directed to the general reader, and all material in his special field including official and professional documents and correspondence; can read reasonably legible handwriting without difficulty.

Native or Bilingual Proficiency

Speaking-5. Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

Reading-5. Reading proficiency equivalent to that of an educated native. Can read extremely difficult and abstract prose, as well as highly colloquial writings and the classic literary forms of the language. With varying degrees of difficulty, can read all normal kinds of handwritten documents.

4. Adequacy of Preparation and Feedback. Standards should be administered with an adequate student advisory and guidance system so that high standards are not accompanied by high failure rates.

--This is an area that will need attention.

Students need to know what will be expected of them. Need ongoing help if they can't meet standard initially. May need help even after leaving school. Could pick up skills and meet standard later.

5. Consistency. Standards should be applied consistently within programs, across programs, and over time.

--This means consistent assessment conditions

- Probably should mean tests with generous time limits or no limit
- "Secure" test forms
- Equating of test forms so a score on a new form will have the same meaning
- Need consistent "exception" procedures.

6. Acceptability/Transferability. Standards should be backed fully by the home institution to increase the likelihood that the resulting diploma will be acceptable to other institutions.

--Need to document and explain your procedures.

Need public relations work here.

--Possibly active local efforts could prevent

state imposed systems. Set your own standards in pilot projects.

~~—May need multiple diplomas or special information on diplomas to satisfy all the different audiences.~~

7. Student Satisfaction. Standards should be administered so that students involved feel they have been well served.

Summary: Standards should be FAIR, CONSISTENT, and REALISTIC

METHODS OF SETTING STANDARDS

Three factors will contribute decisively to the standards setting procedures of any state or school district.

- The external requirements of other institutions, agencies and the like, e.g., state departments of education, state legislatures
- The basic educational mission and objectives of the school district or state
- The personal development goals of individual students

The approach to be used may vary depending on whether only an overall statement regarding reading competency is sought or whether a number of separate decisions about aspects of reading competency will be made.

My discussion assumes that you have a set of reading tasks or questions and want to set a minimum standard (and possibly other higher level standards). In some cases you may also want to have a set of reading objectives, some test questions with each objective, and a standard for each objective. You will still need an overall standard.

Traditional or Arbitrary Method

This method uses some fixed percentage of correct answers on an examination, e.g., 65%, as a passing score or standard. It does not take into account the nature of the skills or content tested, the student population, or other factors. Even though it is the most widely used of any standard setting method, I believe that it has almost nothing to recommend it.

John Geyer -- In New Jersey the Task Force knew standard should be 65% before they were sure what should be measured.

This is arbitrary, other methods not arbitrary in that they involve data collection and systematic judgment done in ways I consider appropriate.

There is a sense in which the other methods can also be considered arbitrary. Reading competence comes in all degrees, and probably varies on many dimensions. Setting a standard divides a continuous distribution.

--Small differences at cut point

--Error of measurement

Review of Test Questions or Tasks

The method involves the use of judges who are familiar with the typical reading performance of students. The judges have to be able to agree, in general terms, on what would constitute minimal reading competency, e.g., how they would recognize a student who was minimally competent or who was not minimally competent. They then need to work with an examination or set of tasks that is already available and which they feel can be reasonably used as a basis for making decisions about whether or not a student is a minimally competent reader. The judges then examine each question or behavioral statement of a critical task and decide if a minimally competent student could answer the question correctly or perform the task adequately. The standard is simply the average number of questions or tasks that each judge has determined should be answered correctly or performed at a specified level of competence by minimally competent students.

Judges can also be asked to state the probability that a minimally competent person would answer each question correctly or they could be asked to

envision a group of minimally competent students and to estimate the portion of that group that would answer each question correctly. In any event, the sum of the proportions represents the minimally acceptable score. Where task review is involved, judges will need to decide on the weight to be given to different aspects of the task, such as the kind and degree of learning reflected in the product or performance.

Use of Preselected or Criterion Groups

This method requires the identification of two groups of students:

- one above minimum competence
- the other below minimum competence

on the particular attribute for which you are setting the standard. You may also want to have a group of students about whom you are doubtful--students that would fall in between the above and the below group. You need to divide the students into the two (or three) groups on some basis other than the examination that you want to use in your operational program. You should use teachers' judgments, course grades, special testing, observations, and self reports. (Use any information that is available or which you can obtain for a reasonable expenditure of money or time.)

It makes sense to devote more time to this activity than you would want to devote in an operational program since this division of the students is going to lead to the setting of a standard on a more practical procedure that can be used on an ongoing basis. After you have your students divided for analysis purposes (you don't have them divided in separate rooms), you are then ready to administer your examination or your assessment procedure to the different groups and obtain a distribution of scores. Your competency standard,

then, is selected to discriminate between the two groups. If your measure does not discriminate it between your two groups (or among your three groups) you should not use it as a basis for setting standards.

Classification Errors

When you are deciding where to set a standard, you have to consider two major types of errors:

- If the standard is set too high, the probability increases that those who should be considered competent will be classified as not competent.
- If the standard is set too low, the probability increases that those who should fail will be passed.

Clearly the setting of standards must take into account the costs or losses associated with misclassifications of each type. In every situation, basic principles of fairness, realism, and consistency will apply.

An observation on this process--the job of setting appropriate performance standards is as critical as it is difficult. Developers of evaluation procedures as part of minimal competency requirements should allow sufficient time in their projects for a thoughtful analysis of the possible approaches. More than one method of setting a standard should then be tried, and the results compared and evaluated. After this evaluation, a decision can be made regarding the method or methods that will be used.