



**CENTER FOR
LABOR MARKET STUDIES**

M E M O R A N D U M

TO: Gail Spangenberg
Council for Advancement of Adult Literacy/
National Commission on Adult Literacy

FROM: Andrew Sum
Director, Center for Labor Market Studies
Northeastern University

DATE: July 23, 2008

RE: Thomas Sticht's Recent Posting About the Response Probability (RP) Standards for Literacy Assessments; My Own Views About the Appropriate Choice of RP Values

In a recent commentary on our report, Thomas Sticht critiqued our interpretation of the NAAL assessment's findings on the literacy proficiencies of adults by pointing out that a placement of a person "below basic" performance level does not mean that he or she cannot do certain tasks but that they only have a less than 67% chance of doing so. In my review of the National Research Council's report on the NAAL Performance Standards, I was very critical of any attempt to use a Response Probability of .50 or .67 as minimally proficient. Only at Harvard University and a few other elite institutions would a 65 % get you an A on your official grade transcript (See: Harvey Mansfield's articles on grade inflation at Harvard).

In my review, I raised the following questions about the choice of an RP standard and listed a set of critiques of the .50 and .67 RP values, which I believe are far too low. Excerpts from my review of the NALS study are attached to the back of this memo. Below I lay out a range of choices we all make in life (from the choice of a restaurant, dentist or doctor to our moral code of conduct) and ask all of you (fellow commissioners) to consider whether you would accept an RP value of .65 or .50 as acceptable in any of these domains. Please pay attention to my ending questions about abiding by the Ten Commandments.

Selecting Minimally Acceptable Performance Standards in Life

- A. Would you go to a dentist that did a root canal on the right tooth 70% of the time? Hey, what's a tooth? We got a ceramic replacement for you.
- B. Would you patronize the local restaurant that got your order right 65% of the time? Or the local bar that brought you the right drink 70% of the time. Sauvignon blanc, when chenin blanc, semi-blanc, what's the difference?
- C. How about the newspaper company that delivered your paper 80% of the time? Nothing new yesterday. Dilbert is still clueless.
- D. Would you want to undergo bypass surgery with a heart doctor that proudly said "75% of his patients survived the operation"? Life ain't everything.
- E. How about being married to a spouse that is faithful 80% of the time? All are welcome to come to the Tom Sticht wedding chapel. Do you take this man to be your lawfully wedded husband and agree to stand by him 80% of the time? Or 70% of the time until death? Loretta Lynn, where are you when we need you? The song's lyrics do not read "Stand by Your Man As Much As You Can for 70% of The Time".
- F. Or when you reach the pearly gates and God asks how often you abided by his commandments on earth, do you think you stand a good chance of being rated as minimally proficient for entry into heaven if you say the following:
- I worshipped only one God about 75% of the time.
 - I only killed twice in my lifetime and they deserved it.
 - I honored my mother and father about 70% of time, when I felt good.
 - I was faithful to my spouse over 80% of the nights we were married. Hey, Lord, who's perfect?
 - I coveted my neighbor's goods only 10% of the time, and I returned the lawn mower before they knew I took it.

I believe that someone who met the above standards will have a lot of time debating RP standards with the devil. God Bless America and save us from the RP 65 and 50 crowd.

The Selection of Response Probabilities for Identifying Item Difficulty and Selecting Cut Scores for Performance Levels on the NAAL

The research report by the National Research Council does a commendable job in explaining the importance of response probabilities (RP values) in analyzing the findings of a literacy assessment such as the 2003 National Assessment of Adult Literacy (NAAL). The selection of RP values will determine item difficulty levels, influence the selection of cut scores for the performance levels, and determine the distribution of the nation's adults across performance levels. The use of alternative RP values (50, 67, 80) in the first NAAL workshop on performance levels and cut scores and the use of only RP 67 values in the second workshop was carefully described and explained. I believe, however, that the discussions of the use of RP 50 values in the re-analysis of the 1992 NALS data, the statistical advantages of RP 50 values, and the potential uses of RP 50 values in setting performance standards are woefully inadequate and misleading. Any use of RP 50 values or even RP 65 values in setting performance standards or determining cut scores should be discarded or seriously revised. The use of any results based on RP 50 values would cast serious doubt on the validity or policy relevance of the NAAL results. Let me explain my case and back my opinion up with the views of my own panel of experts who I surveyed on this issue after completing my review of this study. I did not discuss any of the report's findings with this group but simply asked them what RP values they would desire in a report on literacy that they would receive from a hypothetical literacy assessment in their area.

The choice of an RP 50 or even an RP 65 is, in my opinion, completely indefensible and representative of the dummying down of American educational standards. I would not ever wish to defend such a mediocre standard in any court, public policy seminar, Congressional testimony or newspaper article. An RP 50 or 65 standard at best could be described as low level mediocrity. Sir Max Beerbohm is once reported to have said, "only mediocrity can always be expected to be at its best."

The report does a very nice job showing how the uses of alternative RP criteria will influence the placement of test items along the difficulty scale and the reliability of the scale scores that would be associated with different RP values. It is true that the use of RP 80 values does lead to less statistical precision in the estimates of scale scores than lower RP values, especially RP 50. However, I find that the difference in the confidence intervals of the scale score estimates for RP 50 and RP 80 to be quite small, especially in the context of the overall

degree of variance in scale scores on the past NALS test. A 2 to 3 point difference in the error range of the scale score estimates between RP 50 and RP 80 values amounts to only .03 to .05 of a standard deviation, a relatively small amount of dispersion. This greater degree of statistical precision of estimates based on the RP 50 criteria needs to be judged in the context of the lack of relevance of an RP 50 standard. The set of estimates that has more statistical precision should be valued more highly, *ceteris paribus*. But the standard of mastery is not being held constant, and the huge gap between proficiency at the .50 and .80 level does not make the greater statistical precision of the RP 50 values sufficiently worthwhile to classify it as an acceptable alternative.

Since I have such strong opinions on the desired RP levels for reporting results and establishing cut scores, I solicited the opinion of a sample of individuals across the country who currently work or have worked in the field of education, job training, educational research, and workforce development program planning and evaluation. The professional backgrounds of these 12 individuals are displayed in Table 1. Their educational backgrounds range from high school and apprenticeship in skilled blue-collar occupations to Ph.D.'s in economics and education. I solicited their opinions on three issues. What RP value (50, 65, 80) would they want to use in selecting performance standards and cut scores for literacy tests and in reporting results? Why do they prefer their chosen standard and would they prefer some other value than these three? How would they respond to a seminar presentation on literacy performance if the speaker reported that he used an RP 50 value in generating his findings on performance?

Table 1
Professional Backgrounds of Panel Respondents

1. Vice President of an educational and remediation training company in Virginia that designs curriculum for a diverse array of educational programs, including junior high and high schools, alternative schools, community colleges, and adult basic education programs.
2. Senior research associate (Ph.D. in Economics) in a research center devoted to workforce development and training issues in the New England region.
3. Community college professor (MBA) and Chair, Department of Business, in a community college in Illinois.
4. Director, Local Workforce Development Board in Massachusetts.
5. Director of a large workforce development agency in a large central city in Massachusetts.
6. Director (Ph.D. in education) of a university-based research firm devoted to workforce development issues in Massachusetts.
7. Director of an alternative schools network in the state of Illinois
8. Office manager/trainer for administrative support staff in a manufacturing firm in Indiana.
9. Research director (Ph.D. in Education) for a quasi-public workforce development agency in the state of Massachusetts.
10. Assistant director for a marketing research department with staff training responsibilities in a national financial services firm.
11. Plant foreman for maintenance and installation of capital equipment in a large steel manufacturing facility in the Midwest.
12. Director of a state adult basic education agency in a large Southern state.
13. President, national research corporation in the areas of education and training.

My findings from these interviews are summarized in Table 2. Not one of the 12 panelists (all of whom were interviewed individually either in person or over the phone) would have chosen the RP 50 value. In fact, as a group, they were overwhelmingly opposed to the RP 50 value and felt that its use in developing performance standards or reporting results would be detrimental to any study. The use of RP 50 as a mastery standard was viewed as “laughable”, “meaningless”, and “not even coming close to mediocrity”. A clear majority selected the RP 80 standard, about one-third felt that an RP of .70 to .75 was acceptable, and one of the twelve chose the RP 65 although he felt that it would be wise to compare findings under the RP 65 and RP 80 but not with an RP 50. The integrity of anyone presenting findings with an RP 50 standard would be seriously questioned. The panel was even harsher in their judgment of the RP 50 criteria than I was. I would not recommend any reporting of findings on the RP 50 results although I heartily concur with proposals for presenting findings that will illustrate that a person

with an estimated proficiency lower than the one related to a particular test item can answer that question correctly with some given lower probability. It is crucial to constantly illustrate that literacy is not a dichotomous (yes/no) type of skill as the media would like to present.

Table 2:
Respondents' Choice of An Appropriate Response Probability Value for Selecting
Literacy Performance Levels and Their Reactions to A Hypothetical RP50 Value

Respondent	(A) Preferred RP Value	(B) Rationale For Choice	(C) Reaction to An RP 50 Value
#1	80	Proficiency has to be set at a value that would garner widespread support. 80 is a reasonable standard.	An RP 50 value would set the bar far too low, the wrong signal to our educators
#2	80	I would consider even picking a higher standard; a more rigorous standard like 90 percent	50 as a standard would really be unacceptable. It is like accepting a standard for a coin toss, far too low. I could not imagine reporting results on such a standard to the supervisors of trainees
#3	80	I would prefer a higher standard, but need to be realistic about expectations	An RP of "50" could not meet the face test in a presentation to a workforce development audience. I would expect the audience to laugh.
#4	80	Anything less than 80 would imply a lack of mastery of what is being taught; to me 80 is what is proficient	"I would throw my beverage at the presenter. I would be fired from my job if I performed that poorly on meeting my assigned tasks".
#5	75 – 80	80 may be a little too stringent; 65 is too limited a proficiency and 50 is way too low	I would reply that 50% is not a very valid or meaningful standard; I would not apply this standard to anyone in my office
#6	70 – 75 would be acceptable; 65 is too low	I would accept 70% but would regard it as equivalent to mediocrity	An RP 50 is way too low a standard; I would not take the results seriously from such an assessment. We are making educational standards a laughing stock.
#7	70 percent would be better than the 65 percent as a minimum level of proficiency.	"..for my clients (young welfare mothers) 80 percent might be somewhat too demanding, but 50 percent would represent too low a level of understanding".	50 percent is too low a level of performance to be judged as satisfactory. We would be setting our standards too low and sending the wrong signals to our students and adult learners.

Respondent	(A) Preferred RP Value	(B) Rationale For Choice	(C) Reaction to An RP 50 Value
#8	At least 80 percent	“I am a perfectionist, in my quality control line of work, I would not consider anything as adequate without being totally right	I would be shocked if anyone would use a 50% standard. How could you consider such a low measure as an adequacy measure.”
#9	80	“I might be willing to accept 75 as a compromise standard for measuring competency in many of the adult basic education programs in my company evaluations.”	“I would tell the speaker that he was ‘full of ____,’ 1 of 2 performance is not proficiency. It is no better than a crapshoot. I would never hire any worker with a 50 percent proficiency in his skill area”.
#10	80 at a minimum	“Even 80 is a loose standard for mastery. I would never consider anything below 80 as measuring a proficient level of performance	An RP 50 is far too low a response probability. Any speaker indicating that he used an RP 50 would be written off immediately. 50 is so low that it is meaningless.
#11	An 80 would be my minimum proficiency standard; 65 is too low to be considered acceptable.	It would be useful to compare the distribution of adults by performance level with a 70 and 80 RP.	The RP 50 measure would be a sure fire way to have the literacy assessment dismissed by the public, the business community, and by key educational policymakers.
#12	An RP 65 would be my minimum acceptable standard. I would like to compare findings with the RP 80 results.	Some adult educators felt that the RP 80 level under NALS was too high a standard. Some compromise between 65 and 80 would be ideal	50 is too low a measure. It would cast doubt on the findings of the assessment