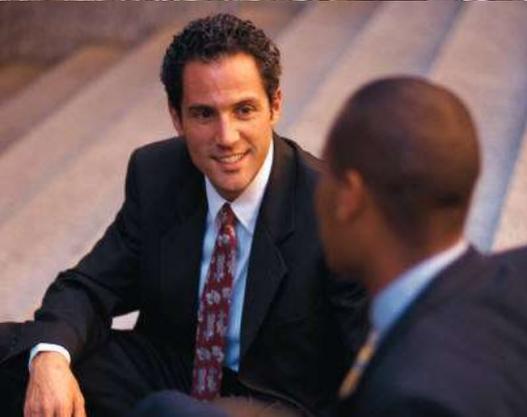




Navigating the Maze of Professional and Legal Guidelines for Setting Cut Scores



Joshua M. Sacco
Aon Consulting
Rick DeShon
Michigan State University



June 27, 2006

Table of Contents

- Background
 - Why are cut scores important?
 - Professional and legal guidelines
 - Why you shouldn't set cut scores
 - When are cut scores most likely to be challenged?
- Minimum Qualifications vs. Business Necessity
- Public Sector vs. Private Sector
- Professional Guidelines
- Legal environment
- Recommendations
- Additional considerations
- Conclusion
- References

Why are cut scores important?

- Candidate flow
 - Cut scores determine the resources that an organization will expend to conduct the testing process in multiple hurdle systems
- Candidate time investments
 - Cut scores determine how much time candidates will invest in the testing process in multiple hurdle systems
- Hiring/Promotion decisions
 - Cut scores will play an important role in determining which candidates are ultimately selected/promoted/certified

Why are cut scores important? (cont'd)

- Plaintiffs alleging illegal discrimination in testing must clear 3 hurdles
 - Adverse impact – Plaintiff Burden
 - Validity – Defendant Burden
 - Equally valid, less adverse alternatives – Plaintiff Burden
- The cut score helps determine the level of adverse impact
- Lowering a cut score can be an equally valid, less adverse alternative
- Cut scores are being litigated at an increasing rate. We expect this trend to continue or accelerate
 - Test users are becoming increasingly sophisticated about using valid tests, especially for large-scale testing
 - The use of valid selection tools is no longer cutting edge
 - Validity itself should be a less enticing target
 - There is a lack of clear and specific guidance on cut scores in the legal and professional literature...more room for interpretation and litigation

Professional and Legal Guidelines

- *SIOP Principles*
 - A litany of issues to consider when setting cut scores
 - “As high or as low to meet the requirements of the organization”
- *APA Standards for Educational and Psychological Measurement*
 - Standard 4.19
 - Document procedures used to set cut scores
 - Ensure SMEs are qualified, adequately trained, given feedback, etc.
 - Adequate measurement precision at regions of the scale where cuts are made
- *EEOC Guidelines*
 - “cutoff scores... should... be reasonable and consistent with normal expectations of acceptable proficiency within the workforce”
 - Ranking on a valid test and using a higher cutoff for applicants who have “little or no chance of being selected may be appropriate, but the degree of adverse impact should be considered”
- *1991 Civil Rights Act*
 - “selection...must bear a significant relationship to successful performance on the job...[and] demonstrable evidence is required”

Why you shouldn't set cut scores

- Cut scores are set to minimize the number of candidates who proceed to a further step in the process
 - Consistent with a non-compensatory approach to selection
- Given unlimited resources, there would be no need for cut scores
 - Minimum qualifications may be an exception in some cases
- Top-down selection, where possible, may be more beneficial than multiple hurdle approaches (i.e., setting cut scores)
 - Information from all predictors is considered
 - It reduces false negatives
 - The number of openings drives who is hired, eliminating the opportunity to litigate the cut score
 - Tradeoffs between resources needed to hire and job performance likely make this a reasonable approach
 - though this is an empirical question that varies by situation
 - Gives all candidates an opportunity to demonstrate all relevant qualifications for a job
 - Generally ignores the continuum of test scores
- When cut scores are set, they should be relatively low, with top-down selection applied in a following step using all the information available

When are cut scores likely to be challenged?

Most Likely	Least Likely
Cut score	No cut score
Highly visible process	Low visibility process
Large organizations	Small organizations
Mass testing	Sporadic testing
High adverse impact	Low/no adverse impact
High cut score	Low cut score
Knowledge test cut very different than 70%	Knowledge test cut set at roughly 70%
History of challenges in the organization	No history of challenges in the organization
Union	No union

Minimum Qualifications vs. Business Necessity

Minimum Qualifications	Business Necessity
Probably the lowest cut score	Depends on circumstances; generally higher than MQ (rarely lower)
Close to objective	Less objective
Consistent with EEOC Guidelines	Consistent with Professional Guidelines
Will resonate with lawyers	Will resonate with business leaders and HR staff
Equally defensible in all situations	Most defensible when high-impact outcomes are at stake
Most defensible	More risky
Fairly consistent over time	Likely to change over time
Resource-intensive to develop	Resources depend on approach

Public vs. Private Sector Selection

Public Sector	Private Sector
Often there are very low selection ratios (e.g., <20%)	Moderate selection ratios
Highly visible process	Low visibility process
Mass testing	Sporadic testing
Frequent use of tests with high adverse impact (job knowledge & physical ability tests)	High adverse impact testing less common
Long history of challenges	Very few challenges
More likely to have public and/or political pressure to be diverse	Less likely to have public or political pressure to be diverse
Occasional use of pre-defined cut scores	Rare use of pre-defined cut scores
Formal appeals processes are common	Appeals processes are rare
Relatively little urgency to fill positions quickly	Often a strong desire to fill positions quickly
Union	No union

Legal environment

- Lanning v. SEPTA (2002)
 - Discrimination must be supported by “business necessity”
 - “minimum qualifications...likely to be able to do the job”
 - Cannot “wholly defer to an employer’s judgment...of what is desirable”
 - Organization has the right to improve workforce
 - “more is better” is not a sustainable defense against a cut score above minimum qualifications
- IBEW v. Mississippi P&L (2006)
 - “specific and sizable savings” and chance of being “above average” versus “bottom-third” of workers is sufficient to demonstrate business necessity
- Lewis v. City of Chicago (2005)
 - “administrative convenience” does not constitute “business necessity”
- US v. Garland (2004)
 - Business necessity can be shown where “selection criteria bear a manifest relationship to the employment in question” as long as the employer’s “legitimate employment needs are significantly served” by the selection practice

Legal Environment (cont'd)

- Isabel v. City of Memphis (2005)
 - Cut scores negotiated before the test is implemented may not be defensible if they don't assess minimum qualifications
 - Other statistical evidence besides the 4/5ths rule will be examined to establish adverse impact (i.e., statistical significance and effect size)
 - Setting a cut score based on a narrow predictor of job performance is not defensible
- US v. Delaware (2004)
 - Tests with low validities will be given close scrutiny
 - Validation studies for discriminatory tests done after the fact in anticipation of litigation will be suspect

Legal Environment - Summary

- Two standards must be met
 - “Job related”, i.e., valid
 - “consistent with business necessity”, i.e., measures minimum qualifications in order to successfully do the job
 - US v. Garland takes a broader view of “business necessity”
 - Not applicable if it is being used as a mere pretext for discrimination
 - Administrative convenience is not business necessity
 - It might be acceptable if the goal is to screen out those who score so low they have only a small chance of being hired – has this been litigated?
- Plaintiff must show an alternative without a discriminatory effect would also comparably serve the employer’s legitimate interests
- Don’t negotiate cut scores *a priori*, set cuts on narrow predictors, or solely use the 4/5th rule to examine adverse impact
- Public vs. Private
 - Strikingly few testing discrimination cases against private employers go to trial or are settled by the EEOC
 - Business necessity is reasonably clear for the public sector, far less so for the private sector, especially if an organization is setting challenging job performance objectives
- Details of the individual case may trump these general guidelines

Recommendations

- First, decide whether you need a cut score
 - Can the entire process be administered to all candidates, using top-down hiring based on the number of available openings?
 - What is the price of administration vs. litigation?
 - Giving all candidates the opportunity to demonstrate a broad range of skills may enhance candidate reactions to the process
 - Will there be contention around the specific cut score selected?
 - Is the nature of the test well disposed to determining cut scores?
 - Behavioral anchors
- Second, determine whether to conceptualize business necessity as minimum qualifications or a higher level of performance due to key strategic needs of the organization
 - Define the standard used
 - Document the procedures used to gather data about the standard
 - Arrive at a quantifiable standard that you are willing/able to defend

Recommendations (cont'd)

- Third, determine the cut score using the standard chosen in Step 2
 - Non-Behavioral assessment, no criterion data
 - Angoff or very similar method...judgmental process applied to predictors
 - Non-Behavioral assessment, with criterion data
 - Regression to link predictor scores to a specific level of job performance
 - Behavioral assessment, no criterion data
 - Angoff-like judgments are implicit in formulating the behaviorally anchored rating scales, i.e., determining which behaviors are labeled as effective vs. ineffective
 - Carefully document the process and SME involvement as discussed earlier
 - Behavioral assessment, with criterion data – tentative
 - Use a combination of Angoff-like data along with regression
- Fourth, determine whether and to what extent cut scores should be communicated

Additional Considerations

- Consider setting the cut score at 1 SE lower than the identified predictor score corresponding to the standard of interest
 - This is a conservative approach that should reduce the likelihood of successful challenge
- Probability of selection as a cutscore designed to save individual and organizational resources
- Consider the importance of the job performance domain being targeted
 - e.g., safety issues will carry significant weight
 - Ensure broad rather than narrow predictors
- Challenges when criterion data exist
 - Performance ratings are often inflated even when gathered for research purposes
 - Many “objective” metrics are subject to a wide variety of factors unrelated to individual performance

Conclusion

- Use only when necessary
- Don't accept the default “70%” solution
- Justify and Document
 - Post-hoc justifications are rarely successful
 - Content and criterion-related validity evidence do not, on their own, justify the use of a cut score...the cut score still must be defensible
 - Administrative convenience (i.e., cost) is not a justification
 - Follow professional standards and use regression-based predictions of acceptable performance on a criterion or one of the variants of the Angoff method
 - Document the consideration of alternatives
- Don't rely exclusively on the 4/5ths rule to evaluate adverse impact
 - Plaintiff will surely look at other indicators of adverse impact...consider the practical size of the subgroup differences
- Hire a good expert and a good attorney
- Keep your fingers crossed!

Reading List

- Albermarle Paper Co. v. Moody, 422 U.S. 405 (1975).
Bew v. City of Chicago, 252 F.3d 891 (7th Cir. 2001).
Dothard v. Rawlinson, 433 U.S. 321 (1977).
Fitzpatrick v. City of Atlanta, 2 F.3d 11 12 (11th Cir. 1993).
Griggs v. Duke Power Company, 401 U.S. 424 (1971).
Gulino v. Board of Education, 96 Civ. 8414 (S.D. N.Y., 2003).
International Brotherhood Of Electrical Workers, V. Mississippi Power & Light, 442 F.3d 313 (5th Cir. 2006)
Isabel v. City of Memphis, 03-5912 (6th Cir. 2005).
Lanning v. Southeastern Pennsylvania Transportation Authority, 181 F.3d 478 (3d Cir. 1999).
Lanning v. Southeastern Pennsylvania Transportation Authority, 2002 U.S. App Lexis 21506 (2002).
Lewis v. City of Chicago, 98 Civ. 5596 (N. D. Ill., 2005).
US. V. Garland, 3:98 Civ. 0307-L (N.D., TX,, 2004).
US v. Delaware, 01 Civ 020-KAJ (DE, 2004).

Reading List (cont'd)

- American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cascio, W. F. & Aguinis, H. (2005). Test development and use: New twists on old questions. *Human Resource Management, 44*, 219-235.
- Kehoe & Olson (2005). Cut scores and employment discrimination litigation. In Landy (Ed). *Employment Discrimination Litigation: Behavioral, Quantitative, and Legal Perspectives*.
- Society for Industrial/Organizational Psychology (2003). Principles for the validation and use of personnel selection procedures (4th Ed). Bowling Green, OH: SIOP.