Supportive Data & Guidelines for Using the

Angoff, Ebel and Nedelsky Cutoff Score Methods

William E. Donnoe
Donnoe & Associates
5743 Marconi Avenue
Carmichael, CA 95608

Roy P. Amato
Superior Court Personnel Services
700 Civic Center Drive West
Santa Ana, CA 92702-1994

Draft copy to be presented at the International Personnel Management Association Assessment Council (IPMAAC) Conference, Newport Beach, June 1997. Not to be cited or quoted without permission from the authors.

Supportive Data & Guidelines for Using the

Angoff, Ebel and Nedelsky Cutoff Score Methods

Abstract

The paper examines and discusses the Angoff, Ebel and Nedelsky cutoff score methods and a customized approach to the Angoff method. Empirical data from two occupational licensing examinations suggests that: (1) the Angoff method is preferable to the Ebel and Nedelsky methods; (2) judgments of individual test items are not affected by a group vs. a one-to-one setting; (3) there is no significant difference between the pass point estimate obtained using the traditional and a customized approach to the Angoff method ($p < .05$); and (4) the statistical similarity between the traditional and the customized approach to the Angoff method is significant ($p < .001$). Implications and limitations of the paper are discussed.

Supportive Data & Guidelines for Using the Angoff, Ebel and Nedelsky Cutoff Score Methods

Introduction

The field of personnel assessment and selection has relied on three content cutoff score methods (i.e., Angoff, Ebel, Nedelsky) to determine the pass point on multiple choice examinations. There are three essential reasons why these content methods have been utilized to set pass points: (1) they assist in identifying individuals who possess the minimum required knowledge, skills and abilities (KSAs) to perform essential job tasks; (2) they assist in maintaining a minimum standard of job related skills and competencies; and (3) they assist in selecting the best when applicants and/or candidates exceed job openings. However, in a demanding work environment, personnel assessment/selection professionals may be reluctant to use these methods given the demands they place on the part of subject matter experts (SMEs). These demands include having SMEs attend a training session in which they review tasks, KSAs required at entry and define minimum acceptable competency. Following such training, the SMEs review, as a group, the examination items and rate the probability that marginal individuals would respond correctly to the items.

Problem Statement

The introduction suggests that while there are benefits to using the Angoff, Ebel and Nedelsky methods to set pass points on multiple choice examinations, they may not be utilized given the demands they place on SMEs. To provide effective services in a demanding work environment, an assessment/selection professional should know which of these methods is best suited for a given examination and if a method may be customized to reduce the demands placed on SMEs. While this information is essential,

there seems to be a gap in the personnel selection literature that documents such information. The paper proposes to fill this gap. In doing so, it will examine empirical data from two occupational licensing examinations to provide personnel assessment/selection professionals significant information and guidelines to use in setting pass points on multiple choice examinations.

The paper will proceed as follows: First, a brief discussion of the Angoff, Ebel and Nedelsky cutoff score methods is presented. The intent of this discussion is to highlight the demands placed on SMEs in setting a cutoff score. This brief discussion is followed by a presentation of the results obtained from examining empirical data from two licensing examinations. The paper concludes with a discussion of the implications and limitations of the findings to setting cutoff scores on multiple choice examinations.

The Angoff. Ebel & Nedelsky Cutoff Score Methods

The standards for Educational and Psychological Testing and the Principles for the Validation and Use of Personnel Selection Procedures are the two principal sets of professional guidelines that support the use of the Angoff, Ebel and Nedelsky methods to set cutoff scores. As set forth in these guidelines, the methods rely on SMEs to work in a group setting (preferably 5 or more SMEs working under the guidance of a facilitator) to appraise the content of individual items or sets of items comprising an examination. After appraising the items, the major task for SMEs is to estimate how examinees will perform on individual items or set of items. Typically, SMEs will define a minimally acceptable candidate (MAC) and then estimate how such a candidate would answer individual items or sets of items. As will be discussed below, the specific tasks of the SMEs and the steps to calculate the cutoff score is what distinguishes one content method from another.

Nedelsky Method

Developed by Leo Nedelsky in 1954, this content method sets "absolute standards" for setting a cutoff score on multiple choice examinations (Meskauskas, 1976). The method is based on the theory that marginal test takers (i.e., test takers who possess relatively low levels of the KSAs tested on an examination) will eliminate as many incorrect choices from an item and then guess from the remaining alternatives (Livingston & Zieky, 1982; Meskauskas, 1976).

This method presumes that minimally acceptable test takers possess sufficient levels of the KSAs tested for and that they are able to eliminate obviously incorrect choices from an item. It is the task of the SMEs to identify for each item on an examination the obviously incorrect choices in an item and then to calculate the expected score for each item. The sum total of the expected scores is the cutoff score or the score that a MAC would be expected to get on the examination.

Angoff Method

The Angoff method is similar to the Nedelsky method in that the cutoff score is calculated using the expected score of each item on the examination. The difference is that the Angoff method does not require SMEs to eliminate the obviously incorrect choices in each item on an examination. Instead, SMEs consider each item on an examination as a whole and make a judgment of the likelihood that a MAC will answer the item correct.

The judgment for each examination item is based on a percentage. The sum total of the percentages is the cutoff score or the score that a MAC would be expected to get on the examination.

Ebel Method

The Ebel method is a variation of the Angoff method in which the task of the SMEs is to classify examination items into a 3X4 table according to difficulty (easy, medium, hard) and relevance (essential, important, acceptable, questionable), and then have the SMEs make judgments (i.e., percentage) that a MAC will answer the items classified correct. The cutoff score is calculated by multiplying the number of questions classified within each category (i.e., difficulty, relevance) by the SMEs' judgment (percentage) that a MAC would get the items classified correct.

As evident in the above brief discussion, each cutoff score method requires specific steps to be completed and these steps place specific demands on the part of SMEs to calculate a cutoff score. The paper will now proceed in discussing the methodology used in setting the cutoff scores on the two licensing examinations.

Methodology

First Licensing Examination

To set the cutoff score on the first occupational licensing examination, 550 SMEs were contacted throughout the State of California. Out of the 550 SMEs contacted, 80 were scheduled and 29 actually participated. See Table 1.

Of the 29 SMEs that participated in setting the cutoff score on the licensing examination, 5 developed a standard MAC definition that was used to set the cutoff score with Angoff, Nedelsky and Ebel methods. For the Angoff method: (1) 7 SMEs using the standardized MAC definition worked in a one-on-one setting with a consultant and generated item ratings; and (2) 5 SMEs using the standardized MAC definition worked in a group setting and generated item ratings. For the Angoff method, the goal was to determine if there is a difference in a cutoff score set in a group or in a one-on-one setting. For the Nedelsky method, 5 SMEs using the standardized MAC

definition worked in a one-on-one setting with a consultant and generated item ratings. Finally, for the Ebel method, 7 SMEs worked in a one-on-one setting with a consultant and generated item ratings. This methodology allowed each cutoff score method to be compared and contrasted while assuring that all SMEs were using the same MAC definition.

Analysis of variance was conducted to determine if: (1) the cutoff score methods (i.e., Angoff, Ebel, Nedelsky) produced significantly different estimates of a cutoff score and (2) if, for the Angoff method, the type of environment (i.e., group vs. one-on-one with a consultant) significantly affects test item judgments.

Finally, correlation analysis was conducted to: (1) determine the extent to which actual items on the licensing examination were judged by the SMEs with any consistency, and (2) to investigate the relationship between the three cutoff score methods and actual item difficulty. For this latter analysis, the difficulty levels of the items on the licensing examination were correlated with SME item ratings for each cutoff score method. This analysis is based on the hypothesis that a relationship should exist between item difficulty and SME item ratings.

Second Licensing Examination

To set the cutoff score on the second licensing examination, a customized and the traditional Angoff method was used. For the customized approach, a standardized MAC definition was developed as part of the job analysis and test specification report for the occupation. During the review of an existing test item bank, 20 SMEs working independently used the standardized MAC definition to generate ratings for individual items. These 20 SMEs rated items within content areas that they were comfortable with, and most skilled with. Consequently, some SMEs rated a small number of items and others rated a larger block of items. Each item that was assembled into the final licensing examination had Angoff ratings from 5 or more individual SMEs. Finally, after the reviews were completed, a 1 00-item multiple choice licensing examination was developed.

For the traditional approach, 10 SMEs using the standardized MAC definition generated ratings for individual items using the traditional Angoff procedures. This methodology allowed the customized and the traditional Angoff method to be compared and contrasted while assuring that all SMEs were using the same MAC definition.

Analysis of variance was conducted to determine if the type of environment (i.e., group vs. one-on-one with a consultant) significantly affects test item judgments. Finally, correlation analysis was conducted to assess the statistical similarity between the traditional and the customized approach to the Angoff method.

Results

First Licensing Examination

For the Angoff method, results revealed that the percentage of items correct that a MAC is likely to obtain on the licensing examination is in the high .40 to low .50. Thus,

based exclusively on the Angoff method, the approximation of the cutoff score on the licensing examination is around 50% correct.

For the Nedelsky method results reveal, with one exception, that the percentage of items correct that a MAC is likely to obtain on the licensing examination is close to .30. Thus, based exclusively on the Nedelsky method, the approximation of the cutoff score on the licensing examination is 33% or about one third of the items correct.

For the Ebel method the results reveal, with some variability, that the percentage of items correct that a MAC is likely to get correct on the licensing examination is close to .60. Thus, based exclusively on the Ebel method, the approximation of the cutoff score on the licensing examination is 60% or about two thirds of the items correct.

The one-way analysis of variance results revealed that the Angoff, Nedelsky and Ebel cutoff score methods yielded a statistically different cutoff score estimate for the licensing examination, $F(2,25) = 27.09$ $p<.01$. A follow-up analysis using the Tukey procedure confirmed that all three methods generated significantly different cutoff score estimates.

The results of the analysis of variance computation to determine if the type of environment (i.e., group vs. one-to-one) significantly affects test item judgments suggest that there is no significant difference in Angoff judgments made in a group or in a one-to-one setting, $F(2, 13) = 3.18$, $p < .05$. This finding suggests, that for the Angoff method, test item judgments are not affected by a group vs. a one-to-one environment in setting a cutoff score.

The results of the correlation analysis suggest that the ratings obtained on the Nedelsky method are unrelated to the results obtained on the Angoff and Ebel

methods, and that the ratings obtained on the Angoff and Ebel methods are related. See Table 2.

The results in Table 2 support the developmental history of the three methods in that the Ebel method was developed as a variation of the Angoff method. Additional correlation analysis suggests that item ratings based on the Nedelsky and Ebel methods do not correlate with item difficulty levels obtained on the licensing examination. However, although only moderately strong, results suggest that item ratings using the Angoff method do correlate with actual item difficulty levels ($r = .37$, $p < .05$). See Table 3.

The findings in Table 3 may be suggesting that the three content methods are measuring different attributes of an item. Further analysis, however, is needed to confirm this suggestion.

Second Licensing Examination

The customized Angoff method generated a cutoff score estimate of 63 while the traditional Angoff method generated a cutoff score estimate of 64. There is no significant difference between the cutoff score estimate generated on the traditional and on the customized approach to the Angoff method ($t = .12$, $df = 107$; $p < .05$). In addition, the statistical similarity between the traditional and the customized approach to the Angoff method is significant ($r = .634$, $p < .001$).

Discussion

The goal of this final section of the paper is to briefly discuss the implications and limitations of the findings to setting cutoff scores on multiple choice examinations. Given the results obtained from analyzing the data from the two licensing examinations, the Angoff method seems preferable to the Ebel and Nedelsky methods. In addition, a valid cutoff score may be set using the Angoff method without having all SMEs present in a group setting when generating item ratings. Thus, for the Angoff method, once the MAC definition has been defined and accepted, SMEs may work in a one-on-one setting with a facilitator to generate item judgments. In addition, the findings identified that for the Angoff method, a relationship between item difficulty level and SME item judgments exists. This is an important finding given that it suggests that the cutoff score obtained with the Angoff method relates to an index of test reliability. Finally, the findings support that the Angoff method may be customized to reduce the demand of needing 5 or more SMEs working in a group setting in order to generate valid and reliable item ratings.

The findings discussed above are quite encouraging to the field of personnel selection. However, there is always the possibility that inherent limitations may have distorted the findings. The major threat to internal validity that may have distorted the results in this study is selection (i.e., initial differences existing in groups prior to a study that may account for the results). For example there was no control group or random assignment of participants to control for selection as an internal threat to validity. Consequently, further research and exploration is needed to ensure the reliability and validity of the discussed findings. Hopefully, this paper has stimulated an interest in such research.

Table 1. SMEs Participating in Setting the Cutoff Score on the First Occupational Licensing Examination.

Location	SMEs Contacted	SMEs Scheduled	SMEs Appeared
Northern CA	250	32	11
Central CA	134	15	7
Southern CA	166	33	11
	550	80	29

Table 2. Consistency of Item Judgments.

	Angoff	Ebel	Nedelsky
Angoff		.57**	.04
Ebel			.11
Nedelsky			

**p < .01

Table 3. Correlation between Cutoff Score Method & Item Difficulty

Method	Correlation
Angoff	.37*
Ebel	.15
Nedelsky	-.09

*p < .05