An Investigation of Judges' Behaviors Within a Procedure for Setting Cut Scores for NOCTI Occupational Competency Examinations

Richard A. Walter
The Pennsylvania State University

Pennsylvania has maintained a nontraditional pathway for the certification of secondary-level vocational teachers since the 1920s. The key that opens the door to that pathway is the verification of subject mastery via (a) documentation of a learning period in the occupation, (b) documentation of related paid work experience beyond the learning period, and (c) successful completion of an occupational competency examination. For many years, those examinations were developed by personnel of the universities engaged in vocational teacher preparation under the auspices of the Pennsylvania Department of Education and a policy committee, the Pennsylvania Occupational Competency Assessment Consortium. The decision to deny or grant admission to a vocational teacher candidate rested upon a norm-referenced cut score. As specified within the Pennsylvania Policy Manual for Administration of The Occupational Competency Assessment Program (Bureau of Vocational Education, 1977):

> The draft test will be duplicated (50 copies) with excess items and administered to
> 10 occupational instructors and/or occupational incumbents and to as many as 50
> graduating secondary students who prepared for that occupation. Initially test
> norms will be based upon the results of testing 10 occupational

teachers/occupational incumbents, but will be updated as data becomes available through actual use with candidates. (p. 19)

These procedures remained operational until 1975, when Pennsylvania joined the Consortium of States that governs the National Occupational Competency Testing Institute (NOCTI).

NOCTI and its governing consortium of states emerged from the expanding nationwide need for vocational teachers during the mid-1960s. Panitz and Olivo (1970) stated, "Two one-day institutes at Rutgers University (1966), attended by representatives of twenty-three (23) states, concluded that the development and implementation of an occupational competency examination program on a nationwide basis would be a more efficient use of personnel and would provide higher quality examinations" (p. 1). Over its 30 years of continuing development, NOCTI has become a leading provider for occupational competency assessments and services (NOCTI, 2004).

By joining NOCTI, Pennsylvania gained the benefits of the national effort to produce quality occupational competency testing instruments for the nontraditional pathway to vocational teacher certification. This change from developing to purchasing examinations also required the Pennsylvania Occupational Competency Consortium members to revise the procedures for establishing the pass/fail cut scores. The procedures were revised to specify the establishment of the cut score for each written and performance test by subtracting two times the Standard Error of Measurement from the national mean score and rounding the results to the nearest whole number.

However, there has been an on-going problem with the traditional approach of setting cut scores for use by personnel of the Pennsylvania Department of Education in the certification of secondary-level vocational instructors. As detailed within Walter and Kapes (2003):

> By relinquishing control of developing, revising, and piloting to establish normative data for the examinations used to certify vocational teachers to NOCTI, members of Pennsylvania's OCA consortium no longer made the decisions about prioritizing the schedule under which those activities took place. Examinations that remained critical elements within Pennsylvania's teacher certification process were frequently appearing at the bottom of the schedule. The situation became exacerbated by a burgeoning market for student tests that consumed NOCTI resources originally devoted to teacher testing. Although on-going discussions produced changes in the schedule of examination development and revision, the piloting of new and revised examinations to establish normative data from which cut scores could be calculated remained a major problem. The NOCTI staff members were encountering major difficulties in conducting the traditional processes for establishing normative data. The result has been extreme delays in making new and revised examinations available for use; in some cases, three to four years. In Pennsylvania, that has dictated a return to the use of oral examinations conducted by a panel of incumbent workers rather than the more preferable written and performance exams, since the process of certifying new teachers cannot be postponed. Therefore, recently the members of Pennsylvania's OCA consortium decided to investigate alternative procedures for establishing cut scores for NOCTI examinations. (p. 28)

The Walter and Kapes (2003) study was undertaken to answer the question posed by the members of the Pennsylvania Occupational Competency Assessment Consortium, "Is there a viable alternative to the traditional methodology used to establish cut scores for NOCTI examinations?" (p. 40). The authors concluded, based upon the results, that the answer to the question was "yes", and proposed several follow-up studies that might be undertaken to expand upon their initial findings. This article provides a discussion of one such follow-up

study, focused upon the behavior of judges within the application of the Nedelsky (1954) methodology to the NOCTI Audio Visual Communications Technology and Quantity Foods experienced worker written examinations to provide answers to two main research questions:

1. Were the members of the panels of judges able to use the filter of a minimally competent worker to eliminate multiple-choice item distracters?
2. To what extent is there a relationship between the judges' predicted scores for a minimally competent worker and their own achieved scores?

Methodology

Selection of the Examinations

As a result of a conversation with NOCTI staff members during which the persistent problem of securing subjects to pilot experienced worker examinations was reemphasized, it was decided to select both the Audio Visual Communications Technology and Quantity Foods written tests for this follow-up study. Both were newly revised versions of existing written tests currently used in Pennsylvania to certify vocational instructors.

Selection of the Judges

As in the pilot study, the selection of the judges to participate in the application of the Nedelsky (1954) method to these two written tests was a crucial step. Considerations that impacted the selection process included (a) the necessity for judges to possess high levels of expertise in their respective occupational areas, (b) the requirement for between 10 and 15 judges for each panel, (c)) the availability of potential judges, and (d) the need for a broad diversity of employment experiences in terms of work assignments and enterprises. Based upon the pilot study results, as well as the need to balance panel size with manageable expenditures, it was decided to select a minimum of 10 judges for each panel.

Potential members of each panel were contacted via telephone to establish their eligibility and willingness to participate, and to provide them with a brief overview of the project. A follow-up was completed with those selected to participate via a letter within which the goals of the project and the logistics for the convening of the panels were detailed. Difficulties in coordinating the selected date for convening the panels with the calendars of potential members led to the decision to confirm 10 judges and one alternate judge for each panel.

Training the Judges

As emphasized by Behuniak, Archambault, and Gamble (1982), and reinforced by the pilot study (Walter & Kapes, 2003), training the judges to insure their informed participation is an essential step in the process. Therefore, the joint convening of the panels for the Audio Visual Communications Technology and Quantity Foods written tests began with an overview of the process through which vocational teachers are certified in Pennsylvania, the critical role NOCTI examinations play within that process, the protocol to be followed when reviewing the written tests, and the intended application of the outcomes produced as a result of their efforts. The panel members were then provided with an eight-item multiple-choice format pretest based upon the online practice test for the written portion of the driver licensing examination developed by the Pennsylvania Department of Transportation (2002). The panel members were asked to adopt the mindset of a minimally competent driver and use that filter to identify and draw a diagonal slash through the letter of each item distracter that such a person should be able to eliminate as a possible correct answer. Subsequent to panel members' individual completion of the pretest, a group discussion was conducted to assess their level of comfort with the process, answer questions, and facilitate the switch from the filter of minimally competent driver to the filter of minimally competent worker for its application to their

respective NOCTI written test.

Application of the Procedure

Each member of the two panels was provided with a copy of either the NOCTI Experienced Worker Audio Visual Communications Technology or the Quantity Foods written test that did not contain any indication of the correct responses. To insure confidentiality and facilitate the analysis of predicted scores with achieved scores, each was requested to write his/her mother's maiden name on the cover of test booklet received. Panel members were then instructed to independently apply the filter of minimally competent worker to the task of identifying and drawing a diagonal slash through the letter representing the alternate response that could be eliminated as the correct response for each item on the test. A reminder to panel members that they were not expected to select the correct answer, rather simply to eliminate nonplausible ones, was included as part of the final instructions. Each member was also instructed to meet with the researcher in an adjacent area once he or she had completed the task.

Subsequent to each panel member's completion of the assigned task, the elapsed time for which ranged between 57 and 145 minutes, he or she moved to an adjacent area to meet with the researcher. During those meetings, each panel member was instructed to now select the correct answer for each item by circling the appropriate letter. Additionally, each was instructed to indicate with a check mark any item about which he or she wished to comment. Then, subsequent to completion of the second task, they were encouraged to provide written comments, on provided composition paper, regarding the items they had check-marked.

Analysis

Step one in the analysis of the data generated by the two panels of judges was the calculation of the reciprocal predicted scores, or predicted item difficulty (p-values), for all items within each written test (Audio Visual - 200 items/Quantity Foods - 199 items) based upon the number of alternatives eliminated by each judge, as indicated by a diagonal slash through the letter representing that alternative within the test booklet. Both tests consisted of four-alternative multiple-choice items. Therefore, the reciprocals were calculated based upon the following formula: (a) no alternative eliminated, $p = .25$; (b) one alternative eliminated, $p = .33$; (c) two alternatives eliminated, $p = .50$; and (d) three alternatives eliminated, $p = 1.00$. The reciprocals were entered into separate Excel spreadsheets to facilitate calculation of the predicted mean score for each item over all judges, the predicted mean score of all items for each judge, and the predicted mean score of all items over all judges for both tests.

Step two in the analysis of the data was the calculation of the scores achieved by each judge. The letters circled on the test booklets, representing the alternative selected as the correct answer, for each item by each judge were transferred to optical scan sheets and scored using the answer keys secured from NOCTI. The scoring results facilitated the calculation of the achieved mean score for each item over all judges, the achieved mean score for each judge, and the achieved mean score of all items over all judges for both tests.

Step three in the analysis of the data was determining the relationships between the predicted scores and the achieved scores for both written tests. This was accomplished by calculating the difference between the predicted and achieved means across all judges over all items, the correlation between the predicted and achieved means across all judges over all items, and the correlation between the predicted item means and the achieved item means across all judges.

Results

Tables 1 and 2 present truncated versions of the predicted item difficulties for the Audio Visual Communications Technology and Quantity Foods written tests, respectively, based upon the judges' decisions as to which alternative responses would be eliminated as distracters by a minimally competent worker. Within each table, the item numbers are displayed in the first column, the item-by-item reciprocals in the middle columns, and the predicted item mean across all judges in the last column. Across the bottom row are displayed the predicted item means over all items for each judge and the mean of means across all judges at the end of the row.

For the Audio Visual Communications Technology written test, the synthetic item difficulty (p-values) determined by each judge range between .25 (difficult) and 1.00 (easy). The predicted item means for each judge over all 200 items range from a low of .52 to a high of .88, and the predicted item means across all judges range from a low of .28 to a high of 1.00. The overall synthetic mean difficulty of the Audio Visual written test is presented as the mean of means at the right end of the bottom row (.6672). Transformed into a percentage, the theoretical cut score for this test is 66.72%.

For the Quantity Foods written test, the synthetic item difficulty (p-values) determined by each judge also range between .25 (difficult) and 1.00 (easy). The predicted item means for each judge over all 199 items range from a low of .32 to a high of .89, and the predicted item means across all judges range from a low of .40 to a high of .95. The overall synthetic mean difficulty of the Quantity Foods written test is presented as the mean of means at

Table 1
Item Difficulties and Predicted Means for the NOCTI Audio Visual Communications Technology Written Test

Item number	Judge 1	Judge 2	Judge 3	...	Judge 11	Predicted mean
1	1.00	0.33	0.33	…	0.33	0.536
2	0.50	0.25	0.25	…	0.50	0.553
3	1.00	0.50	0.25	…	0.33	0.567
4	1.00	1.00	1.00	…	0.50	0.689
5	1.00	0.50	0.25	…	0.50	0.575
6	0.33	1.00	1.00	…	1.00	0.848
7	0.33	1.00	1.00	…	0.33	0.787
8	1.00	0.50	0.50	…	1.00	0.567
9	0.50	0.33	0.50	…	0.50	0.605
10	0.50	1.00	1.00	…	0.50	0.613
11	1.00	1.00	1.00	…	1.00	0.894
12	0.50	1.00	0.50	…	0.50	0.605
13	0.50	1.00	0.25	…	0.50	0.568
14	1.00	1.00	0.25	…	1.00	0.530
15	1.00	1.00	0.33	…	1.00	0.780

	Judge 1	Judge 2	Judge 3	...	Judge 10	Predicted mean
16	0.50	1.00	0.25	…	1.00	0.461
17	0.50	0.50	0.50	…	1.00	0.682
18	1.00	1.00	0.33	…	1.00	0.643
19	0.50	0.50	1.00	…	1.00	0.727
20	1.00	1.00	1.00	…	1.00	1.000
21	1.00	0.50	1.00	…	1.00	0.886
22	0.50	1.00	0.25	…	1.00	0.583
23	1.00	1.00	1.00	…	0.33	0.666
24	0.25	1.00	0.25	…	0.33	0.408
25	0.50	1.00	0.50	…	0.33	0.491
.	.	.	.	.	.	.
.	.	.	.	.	.	.
195	1.00	1.00	1.00		1.00	0.787
196	0.25	0.25	0.25		0.25	0.386
197	0.25	0.25	1.00		0.25	0.455
198	0.50	1.00	0.50		0.50	0.492
199	0.25	0.25	1.00		0.50	0.598
200	1.00	1.00	0.33		1.00	0.674
	0.63	0.84	0.74		0.78	0.6672

Table 2
Item Difficulties and Predicted Means for the NOCTI Quantity Foods Written Test

Item number	Judge 1	Judge 2	Judge 3	...	Judge 10	Predicted mean
1	0.50	1.00	0.50	…	0.50	0.683
2	0.25	1.00	0.50	…	1.00	0.675
3	1.00	1.00	0.25	…	1.00	0.825
4	0.50	1.00	0.33	…	1.00	0.616
5	0.25	1.00	0.50	…	1.00	0.708
6	0.25	1.00	0.25	…	0.50	0.658
7	0.25	1.00	0.25	…	1.00	0.733
8	1.00	1.00	1.00	…	0.50	0.775
9	0.50	1.00	0.33	…	0.33	0.515

10	0.25	1.00	1.00	…	1.00	0.825
11	0.50	1.00	1.00	…	0.50	0.658
12	0.50	1.00	1.00	…	0.33	0.633
13	0.25	1.00	0.25	…	0.50	0.525
14	0.25	1.00	0.25	…	0.50	0.483
15	0.33	1.00	0.25	…	0.25	0.633
16	0.33	1.00	0.33	…	0.50	0.532
17	0.25	1.00	0.25	…	1.00	0.641
18	0.50	1.00	0.50	…	0.33	0.549
19	1.00	1.00	0.25	…	0.25	0.683
20	0.50	1.00	0.25	…	0.33	0.591
21	1.00	1.00	0.50	…	1.00	0.725
22	0.33	1.00	1.00	…	1.00	0.749
23	0.50	1.00	0.25	…	0.50	0.525
24	1.00	1.00	0.50	…	1.00	0.750
25	0.25	1.00	0.33	…	10.50	0.524
.	.	.	.	.	.	.
.	.	.	.	.	.	.
195	0.33	0.50	0.33	…	1.00	0.541
196	0.33	0.50	0.25	…	0.50	0.441
197	0.25	0.50	0.25	…	0.50	0.433
198	0.50	1.00	0.33	…	0.50	0.666
199	0.50	1.00	0.50	…	0.50	0.750
	0.53	.081	0.44	…	0.57	0.6370

the right end of the bottom row (.6370). Transformed into a percentage, the theoretical cut score for this test is 63.70%.

Table 3 presents the predicted (Mp) and achieved (Ma) means for each judge across all items, the mean of means across all judges for Mp and Ma, the differences within the two sets of predicted and achieved means of means, and the correlations within the two sets of predicted and achieved means of means for the Audio Visual and Quantity Foods written tests.

Table 3
Predicted and Achieved Means, Differences, and Correlations for the NOCTI Written Tests

Judge Foods	Audio Visual Technology		Quantity Foods	
	Pred M_p	Ach M_a	Pred M_p	Ach M_a

1	.63	.69	.53	.73
2	.84	.80	.81	.66
3	.74	.76	.44	.76
4	.52	.73	.76	.67
5	.88	.76	.89	.59
6	.52	.76	.80	.74
7	.60	.80	.54	.69
8	.53	.72	.72	.70
9	.57	.75	.32	.73
10	.73	.62	.57	.78
11	.78	.72	N/A	N/A
Mean of Means	.667	.737	.638	.705
Difference $M_a - M_p$	.070		.067	
Correlation $M_p M_a$	.0653		-.6584	

The ranges of the 11 judges' predicted and achieved means for the Audio Visual Communications Technology written test were .52 to .88 and .62 to 80, respectively, and resulted in mean of means values of .667 (66.70%) and .737 (73.70%), respectively. The difference between the achieved and predicted means of means was .07 (7.00%). The correlation between the predicted and achieved means of means was a negligible value of .0653.

The ranges of the 10 judges' predicted and achieved means for the Quantity Foods written test were .32 to .89 and .59 to .78, respectively, and resulted in mean of means values of .638 (63.80%) and .705 (70.50%), respectively. The difference between the achieved and predicted means of means is .067 (6.70%). The correlation between the predicted and achieved means of means is a moderately strong value of -.6584.

Table 4 presents a truncated version of the 11 judges' p-value decisions, the predicted and achieved item means, and the correlation of the predicted (Mp) and achieved (Ma) item means across all items for all judges on the Audio Visual Communications Technology written test. The correlation between 200 predicted and achieved item means is a moderately strong value of .445. Table 5 presents a truncated version of the 10 judges' p-value decisions, the predicted and achieved item means, and the correlation of the predicted (Mp) and achieved (Ma) item means across all items for all judges on the Quantity Foods written test. The correlation between 199 predicted and achieved item means is a moderately strong value of .511.

Discussion

Based upon the results of this study, it was concluded that the members of the panel of judges were able to use the filter of a minimally competent worker to eliminate multiple-choice item distracters. The findings also noted a moderate positive relationship indicating a lesser expectation for the score achieved by a minimally competent worker.

Adoption of Mindset

The necessity of providing training for the members of the panel of judges to sensitize them to the process was well-documented throughout the literature reviewed (Walter & Kapes, 2003). The validity of this point was confirmed qualitatively

---

Table 4

Correlation of Predicted and Achieved Item Means on the NOCTI Audio Visual Communications Technology Written Test

Item number	Judge 1	Judge 2	Judge 3	...	Judge 11	Predicted mean	Achieved mean
1	1.00	0.33	0.33	…	0.33	0.536	0.640
2	0.50	0.25	0.25	…	0.50	0.553	0.730
3	1.00	0.50	0.25	…	0.33	0.567	0.820
4	1.00	1.00	1.00	…	0.50	0.689	0.730
5	1.00	0.50	0.25	…	0.50	0.575	0.550
6	0.33	1.00	1.00	…	1.00	0.848	0.910
7	0.33	1.00	1.00	…	0.33	0.787	1.000
8	1.00	0.50	0.50	…	1.00	0.567	1.000
9	0.50	0.33	0.50	…	0.50	0.605	0.900
10	0.50	1.00	1.00	…	0.50	0.613	0.450
11	1.00	1.00	1.00	…	1.00	0.894	1.000
12	0.50	1.00	0.50	…	0.50	0.605	0.000
13	0.50	1.00	0.25	…	0.50	0.568	0.180
14	1.00	1.00	0.25	…	1.00	0.530	0.900
15	1.00	1.00	0.33	…	1.00	0.780	0.820
16	0.50	1.00	0.25	…	1.00	0.461	0.900
17	0.50	0.50	0.50	…	1.00	0.682	1.000
18	1.00	1.00	0.33	…	1.00	0.643	0.910
19	0.50	0.50	1.00	…	1.00	0.727	0.820
20	1.00	1.00	1.00	…	1.00	1.000	1.000
21	1.00	0.50	1.00	…	1.00	0.886	0.820
22	0.50	1.00	0.25	…	1.00	0.583	0.910
23	1.00	1.00	1.00	…	0.33	0.666	0.640
24	0.25	1.00	0.25	…	0.33	0.408	0.360
25	0.50	1.00	0.50	…	0.33	0.491	0.450
.	.	.	.	.	.	.	.

	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
1951	.00	1.00	1.00	…	1.00	0.787	0.000
196	0.25	0.25	0.25	…	0.25	0.386	0.640
197	0.25	0.25	1.00	…	0.25	0.455	1.000
198	0.50	1.00	0.50	…	0.50	0.492	0.640
199	0.25	0.25	1.00	…	0.50	0.598	0.730
200	1.00	1.00	0.33	…	1.00	0.674	0.910

Correlation Mp Ma = .445

Table 5
Correlation of Predicted and Achieved Item Means for the NOCTI Quantity Foods Written Test

Item number	Judge 1	Judge 2	Judge 3	...	Judge 10	Predicted mean	Achieved mean
1	0.50	1.00	0.50	…	0.50	0.683	1.000
2	0.25	1.00	0.50	…	1.00	0.675	1.000
3	1.00	1.00	0.25	…	1.00	0.825	1.000
4	0.50	1.00	0.33	…	1.00	0.616	0.800
5	0.25	1.00	0.50	…	1.00	0.708	0.700
6	0.25	1.00	0.25	…	0.50	0.658	0.900
7	0.25	1.00	0.25	…	1.00	0.733	0.800
8	1.00	1.00	1.00	…	0.50	0.775	1.000
9	0.50	1.00	0.33	…	0.33	0.515	1.000
10	0.25	1.00	1.00	…	1.00	0.825	1.000
11	0.50	1.00	1.00	…	0.50	0.658	1.000
12	0.50	1.00	1.00	…	0.33	0.633	0.000
13	0.25	1.00	0.25	…	0.50	0.525	0.100
14	0.25	1.00	0.25	…	0.50	0.483	0.100
15	0.33	1.00	0.25	…	0.25	0.633	0.400
16	0.33	1.00	0.33	…	0.50	0.532	0.600
17	0.25	1.00	0.25	…	1.00	0.641	0.700
18	0.50	1.00	0.50	…	0.33	0.549	0.400
19	1.00	1.00	0.25	…	0.25	0.683	0.600

20	0.50	1.00	0.25	…	0.33	0.591	0.400
21	1.00	1.00	0.50	…	1.00	0.725	0.600
22	0.33	1.00	1.00	…	1.00	0.749	0.900
23	0.50	1.00	0.25	…	0.50	0.525	0.800
24	1.00	1.00	0.50	…	1.00	0.750	0.800
25	0.25	1.00	0.33	…	0.50	0.524	0.900
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
195	0.33	0.50	0.33	…	1.00	0.541	0.700
196	0.33	0.50	0.25	…	0.50	0.441	0.600
197	0.25	0.50	0.25	…	0.50	0.433	0.100
198	0.50	1.00	0.33	…	0.50	0.666	0.800
199	0.50	1.00	0.50	…	0.50	0.750	0.900

Correlation $M_p M_a = .445$

during the training activities by the marked changes in the questions posed by panel members, as well as the shift in attitudes toward the task as expressed through their body language, and quantitatively through examination of the predicted and achieved score data.

Upon arrival, most of the panel members expressed their pleasure at having been invited to participate based upon their occupational expertise. Despite having previously received an overview of the process, most asked a light-hearted version of the same question, "What are we going to do today?" Throughout the introductory presentation on the process of vocational teacher certification, the role occupational competency assessment plays within that process and the necessity of adopting the mindset of a minimally competent worker, the questions posed by panel members became increasingly focused on the specifics and significance of the task. Expression of their attitudes, both verbal and nonverbal, shifted from mild curiosity to intense concentration and even a bit of anxiety. Those changes continued in the same direction as the training progressed through the pretest phase, with the exception of the anxiety on the part of several panel members. Completion of the pretest and the subsequent group discussion of the process resulted in both verbal and nonverbal expressions of confidence in completing the task by the entire group. The veracity of that confidence in their ability to apply the mindset of a minimally competent worker is reflected in the difference between the achieved and predicted means of means. The nearly identical difference values of .070 (7%)for the Audio Visual Communications Technology test and .067 (6.7%) for the Quantity Foods test indicate that, overall, both panels of judges were able to establish a theoretical cut score that is lower than their own level of expertise, as measured by the respective test.

Relationship Between Predicted and Achieved Scores

To further explore the behaviors of judges in this application of the Nedelsky (1954) method, correlation analyses examined the relationships between predicted scores for the minimally competent worker and the scores achieved by the panel members. Expectations were that the analysis would result in positive correlations, thereby indicating that the judges achieved a higher score than they predicted for the minimally competent worker.

The first such analysis was performed on the overall predicted and achieved mean scores. The correlation between the predicted and achieved mean scores (.0653) for the 11 judges assigned to the Audio Visual Communications Technology test was negligible, but in the expected direction. However, the correlation between the predicted and achieved scores (-.6584) for the 10 judges assigned to the Quantity Foods test was moderately strong and in the opposite direction.

A closer examination of the item p-values and achieved means produced a probable explanation of the negligible positive and moderately strong negative correlations. For some items, the judges simply disagreed with the correct answer as designated within the key supplied by NOCTI. Items 12 and 195 on the Audio Visual Communications Technology test, and Item 12 on the Quantity Foods test, provided evidence to support this explanation. The judges awarded each of these items p-values and predicted means that rated them as relatively easy. However, none of the judges selected the correct answer, as indicated by the 0.000 in the achieved mean columns. Further evidence to support this explanation was provided by a review of the written comments about specific test items provided by the examiners subsequent to their analysis and completion of the tests. The majority of their critical comments were directed at the same test items.

The second correlation analysis was performed using the predicted and achieved item means across all judges. The correlation between the predicted and achieved item means for the Audio Visual Communications Technology (.445) and Quantity Foods (.511) tests were both moderately strong and in the expected direction. Clearly, on an item-by-item basis, the members of the panels of judges produced a related overall lesser expectation of performance for the minimally competent worker.

In summary, the underlying assumption of the Nedelsky (1954) methodology is that the judges selected for the panel must be able to understand and apply the concept of minimal competence. These qualitative and quantitative findings confirm the ability of judges to adopt the requisite mindset of a minimally competent worker and apply it to NOCTI written tests.

The findings also support the utility of using judges to establish theoretical cut scores for use in the occupational competency assessment of vocational teacher candidates, provided that the panels are of sufficient size to provide the diversity of p-values required for a valid outcome. Based upon the pilot study and this study, the minimum acceptable size appears to be 10 members.

Recommendations

This follow-up study, based upon the Walter and Kapes (2003) pilot study, was conducted to extend the initial investigation of the viability of an alternate methodology for establishing cut scores for occupational competency examinations. The findings lead to the following recommendations.

1. Members of the NOCTI staff should investigate the feasibility of applying the Nedelsky (1954) methodology to the establishment of initial cut scores for new and revised written tests. Adoption of this methodology would shorten the time lag that currently exists between the development/revision and availability of a test for client use as a result of the difficulties associated with securing an adequate sample to conduct the traditional piloting and normative processes. As discussed in the article detailing the pilot study (Walter & Kapes, 2003), the theoretical scores produced through this methodology may be adjusted through a variety of techniques to establish actual cut scores suitable for the needs of individual NOCTI customers.

2. If NOCTI staff members choose to implement this process, the more traditional normative cut score data should continue to be calculated for use by members of the consortium. This would also facilitate a follow-up study focused on a comparison of the

cut score established via the Nedelsky (1954) methodology with a norm-referenced cut score established for the same written test.
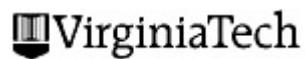
References

Behuniak, P., Jr., Archambault, F. X., & Gamble, R. K. (1982). Angoff and Nedelsky standard setting procedures: Implications for the validity of proficiency test score interpretation. *Educational and Psychological Measurement, 10*, 95-105.

Bureau of Vocational Education. (1977). *Pennsylvania policy manual for administration of the occupational competency assessment program for vocational instructional certification candidates and vocational intern candidates*. Harrisburg: Pennsylvania Department of Education.

National Occupational Competency Testing Institute (NOCTI). Retrieved May 19, 2004, from: http://www.nocti.org

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational Psychological Measurement, 14,* 3-19.

Panitz, A., & Olivo, C.T. (1970). *National occupational competency testing project: The state of the art of occupational competency testing*. New Brunswick: Department of Vocational-Technical Education, Rutgers University.

Pennsylvania Department of Transportation. (2002). *Crossroads: Stories about teen driving*. Retrieved February 20, 2002, From http://www.dmv.state.pa.us/crossroads/quizzes/quizhome.html
Note: The website provided could not be accessed. The link above goes to http://www.dot4.state.pa.us/crossroads_textonly/quizhome.shtml .

Walter, R. A., & Kapes, J. T. (2003). Development of a procedure for establishing occupational examination cut scores: A NOCTI example. *Journal of Industrial Teacher Education, 40* (2), 25-45.

_____

Walter is Associate Professor in the Department of Workforce Education and Development at The Pennsylvania State University in University Park, Pennsylvania, and can be reached at raw18@psu.edu.

DLA Ejournal Home | JITE Home | Table of Contents for this issue | Search JITE and other ejournals