

Adjusting the Passing Scores for Gearing Up for Safety: Production Agriculture Safety Training for Youth Curriculum Test Instruments

William Brian Hoover, Assistant Professor
Murray State University

Brian F. French, Associate Professor
Washington State University

William E. Field, Professor

Roger L. Tormoehlen, Professor
Purdue University

Minimum passing scores for the Gearing Up for Safety: Production Agriculture Safety Training for Youth curriculum (Gearing Up for Safety) were set in 2006 with widely used and established procedures by efforts of subject matter experts (French, Breidenbach et al., 2007; French, Field, and Tormoehlen, 2006, 2007). While providing a research-based curriculum designed to assist in identifying youth who are ready to operate agricultural tractors and machinery, the Gearing Up for Safety curriculum also meets current requirements of the Agricultural Hazardous Occupation Orders (AgHOs) (Ortega, 2003). However, no work has been conducted to re-examine if the evaluative scores are appropriate or if adjustments should be made after field administration of the tests. The purpose of this study was to examine the appropriateness of those passing scores utilizing a compromise or adjustment method (Beuk, 1984). Although the group of experts rejected the recommendations for adjusting passing scores, the panel and researchers were able to review the empirical performance of the three test instruments as a curriculum package for the first time since the Gearing Up for Safety curriculum was developed over a decade ago. Passing standards, established from previous research, were also strengthened through research and re-evaluation (French, Field, et al., 2006, 2007).

Keywords: safety education; certification; validation; farm safety; passing scores

Introduction

Established methods for setting standards (passing scores) of exams for programs that provide certification, placement, or diplomas have been used for decades (Cizek, 1993; Mills & Melican, 1988). These methods mainly involve using expert review and guidance to establish minimum passing scores (Mills & Melican, 1988). However, in many cases, little consideration is given to revision or adjustment of these criteria once established (Mills & Melican, 1988). Employing techniques to examine differences between established passing

scores, set by subject matter experts, and empirical data, gathered from administration of the tests, can provide increased accuracy of the testing and decision process and merits attention by test developers and test program managers (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 1999; Beuk, 1984; De Gruijter, 1985; Geisinger, 1991; Mills & Melican, 1988; O'Neill, Marks, & Reynolds, 2005; Wendt & Kenny, 2007). The results can assist in providing necessary validity evidence for scores to justify inferences and decisions

based on those scores in many types of training or certification exams. Programs such as nursing certification do revisit their standards for passing (e.g., Wendt & Kenny, 2007). However, in other areas such as professional certification programs (engineering, teaching, achievement testing), there is a lack of documentation as to how and on what review schedule such work should be conducted. The straightforward nature of the method employed here may be a valuable tool for other sectors wishing to continue to gather validity evidence for the inferences based on their testing programs.

Passing scores are typically judgments and decisions made by a group of subject matter experts. Adjustments to established scoring criteria should be a judgment of experts as well, even if no change occurs. Minimum passing scores for the Gearing Up for Safety: Production Agriculture Safety Training for Youth curriculum (Gearing Up for Safety) were set with widely used and established procedures through efforts of subject matter experts (French, Breidenbach, Field, & Tormoehlen, 2007; French, Field, & Tormoehlen, 2006, 2007). While providing a research-based curriculum designed to assist in identifying youth who are ready to operate agricultural tractors and machinery, the Gearing Up for Safety curriculum also meets current requirements of the Agricultural Hazardous Occupation Orders (AgHOs) (Ortega, 2003). The AgHOs establish national restrictions concerning employment of youth to work on non-family farms, but are vague concerning methods to adequately assess youth readiness, including minimum curriculum content, test construction, and passing requirements (French, Field et al., 2006, 2007; U.S. Department of Labor, 1971, 2007).

During development stages of the Gearing Up for Safety curriculum, desired core competencies, curriculum content including new subject matter not addressed by the 48-year-old law, an assessment process, and minimum passing standards were established (French, Breidenbach et al., 2007; French, Field et al., 2006, 2007). The curriculum's assessment process was designed to include three major stages in order to evaluate both cognitive and

hands-on abilities to perform certain agricultural tasks considered hazardous (French, Breidenbach et al., 2007; French, Field et al., 2006, 2007; Ortega, 2003). The first stage of the assessment process culminates with a 70 question written exam (Written Exam) designed to evaluate the participants' level of knowledge in the areas of general farm safety, machine safety, and tractor safety. The internal consistency reliability of the exam scores has been high across studies (e.g., 0.87, 0.93). The second stage concludes with a pre-operational exam (Pre-Operational Exam) that evaluates participants' knowledge of basic tractor components and ability to perform basic pre-operational checks of key components and systems. The final stage of testing is an operational exam (Operational Exam) that requires participants to operate a tractor and two-wheel trailer through a standard tractor driving course. Successful completion of the three part assessment process allows participants to be certified under the provisions of the AgHOs (French, Breidenbach et al., 2007; French, Field et al., 2006, 2007; Ortega, 2003).

Developers of the Gearing Up for Safety exams used a group of subject matter experts to set passing standards in 2006. However, there was not a follow-up on these passing scores until now due to a need to allow the process to be in place for a period of time to collect the needed performance data. Given data are now available on actual examinee performance, the purpose of this study was to determine whether the passing scores for each of the Gearing Up for Safety exams needed to be adjusted to meet the demands of the program. Specifically, this study examined the passing scores for all three of the Gearing Up for Safety test instruments through a 'compromise' adjustment procedure (Beuk, 1984). The essence of this method is to compare the set passing scores with scores from completed exams for examinees seeking certification. This comparison helps to determine if the passing rate is in accord with what is expected given the set standard by the panel of experts. Results of the procedure provide data that the expert panel uses to evaluate and discuss if an adjustment to the standard is needed. This method has

successfully been applied to other areas where certification is required (e.g., nursing, Wendt & Kenny, 2007). A secondary purpose of this study was to provide an example of a method that could be employed in other areas of agriculture or other sectors that may require certification. Continually gathering validity evidence for inferences based on test scores helps to ensure good, accurate decisions about individuals.

Procedures

Development stages of the Gearing Up for Safety curriculum utilized subject matter experts in the areas of agricultural safety and health from academic, professional, industrial, and private entities (Bullock, 2006; French, Breidenbach et al., 2007; French, Field et al., 2006, 2007; Kingman, Yoder, Hodge, Ortega, and Field, 2005; Ortega, 2003). Input from these professionals produced non-arbitrary passing standards and strengthened the foundation on which to build the remainder of the program (French, Breidenbach et al., 2007; French, Field et al., 2006, 2007; Kingman et al., 2005; Ortega, 2003). These experts also provided an important element in the adjustment process employed to review the appropriateness of the passing scores due to their ability to make informed decisions about youth and tractor safety (Beuk, 1984).

The adjustment procedure involved four main steps to graphically plot the data and perform the adjustment:

1. Collect, average, and plot mean values of passing scores and rates set by subject matter experts.
2. Plot a line using standard deviations of the values collected in Step 1.
3. Plot examinee success rates on the exam based on standards set by subject matter experts.
4. Perform adjustment.

Each of these steps is discussed in greater detail in the results section. To illustrate the results of each step, tables provide the summary information for each test instrument. However, as the procedure was identical for each test instrument, only a figure representing each step of the Written Exam is presented. The resulting graphs for all three exams, following completion of all four steps, are presented at the end of the results section.

Data were collected from subject matter experts in two separate meetings, each lasting approximately 1.5 days. During the first meeting, a standard setting session was conducted for the Gearing Up for Safety exams. Panel members' opinions of the passing score for all three exams were collected and are summarized in Table 1 as a mean (*k*) for each (French, Field et al., 2006, 2007). During the second meeting, panel members were asked to provide ideal passing rates (found in Table 1 as mean *v*) for each of the three exams based on previously established passing scores. The panel was not informed of any empirical testing results before this data was gathered as per recommendations (Beuk, 1984).

Table 1
Mean Values of Passing Score and Passing Rate Percentages Provided by Subject Matter Experts

	Written Exam	Pre-Operational Exam	Operational Exam
<i>k</i> (passing score)	70%	85%	< 20 infractions
<i>v</i> (passing rate)	82%	84%	82%

Participants for adjustment procedure

In accord with the *Standards for Educational and Psychological Testing*, the expert panel consisted of members who represented many years of experience in this area and had a variety of backgrounds that

represented the farm safety field in general. The complete panel consisted of 16 members. Across rating sessions, the number dropped to 13 due to travel schedules or events beyond the committee's control. However, the expertise represented on the panel was not compromised.

The panel was comprised of 6 professors in the safety and agriculture area, 2 insurance specialists, 2 extension representatives, 2 professionals who conduct agricultural safety programs for youth, 2 professionals who publish safety materials for training, and 2 parents of youth receiving training. As evidenced by panel membership, diverse views that captured the various stakes involved in youth safety in agricultural environments were represented.

Participants for Plotting Examinee Success Rate

Empirical data needed for plotting the examinee success rate were collected in 2008–09 while gathering evidence for validity with each of the three tests. Examinees in the study were from Indiana, Illinois, Kentucky, Tennessee, and Washington. The population consisted of 455 high school and 4–H students where 292 (81.3%) of the students were in grades 9–11. There were 179 (48.1%) 14 or 15 years of age, which is the age range for AgHOs

exemptions, 271 (73.4%) males, 98 (26.6%) females, and 338 (94.5%) were of Caucasian ethnicity.

Results

Figures 1 – 4 below represent the four steps outlined previously in the procedure section. These four figures represent the procedure and results for the Written Exam only. Explanations for specific steps and respective figures will help guide the reader through what was conducted to arrive at the adjustment.

According to standards set by the subject matter experts, 82% of examinees should successfully complete the Written Exam with the passing score set at 70%. These data, collected from the panel, are graphically represented in Figure 1 at point *M*.

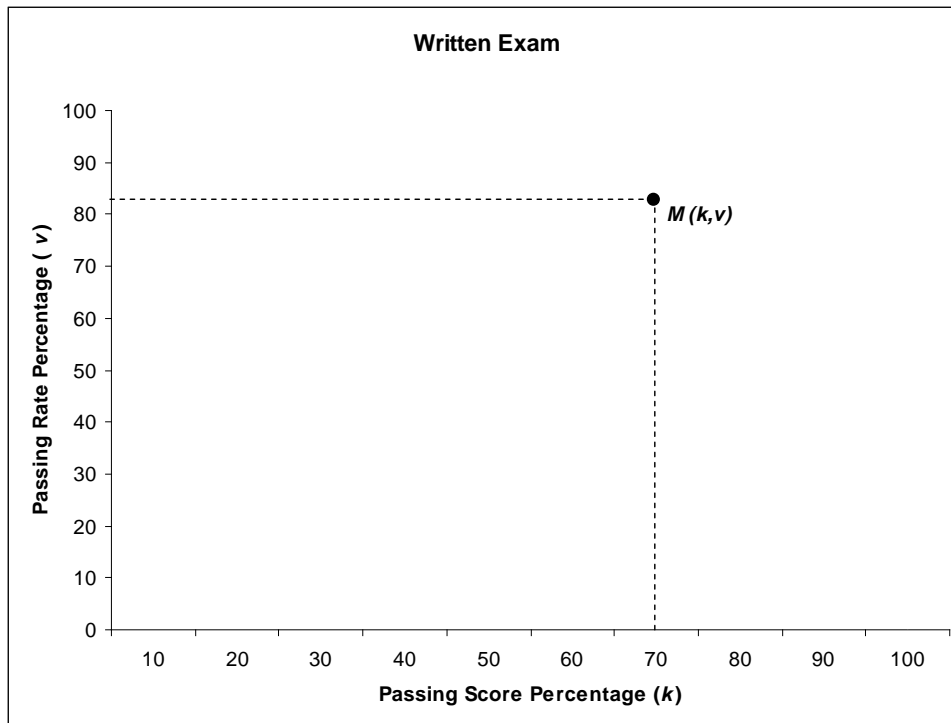


Figure 1. Plot of passing score and passing rate for the Written Exam at point *M* (*k*, *v*).

After summarizing the average passing scores (k) and rates (v) from panel members, the standard deviations for both the passing score and passing rate were determined, s_k and s_v respectively, from the same data set, per recommendations (Beuk, 1984). See Table 2 for

these values with each exam. Standard deviation values were used to determine the slope of a line, with slope s_v/s_k . In Figure 2 the line with positive slope was plotted in addition to point M from Step 1 (Beuk, 1984).

Table 2
Standard Deviation Values of Passing Score and Passing Rates Provided by Subject Matter Experts

	Written Exam	Pre-Op Exam	Operational Exam
s_k	8	4	9
s_v	12	13	12
Slope (s_v/s_k)	12/8	13/4	12/9

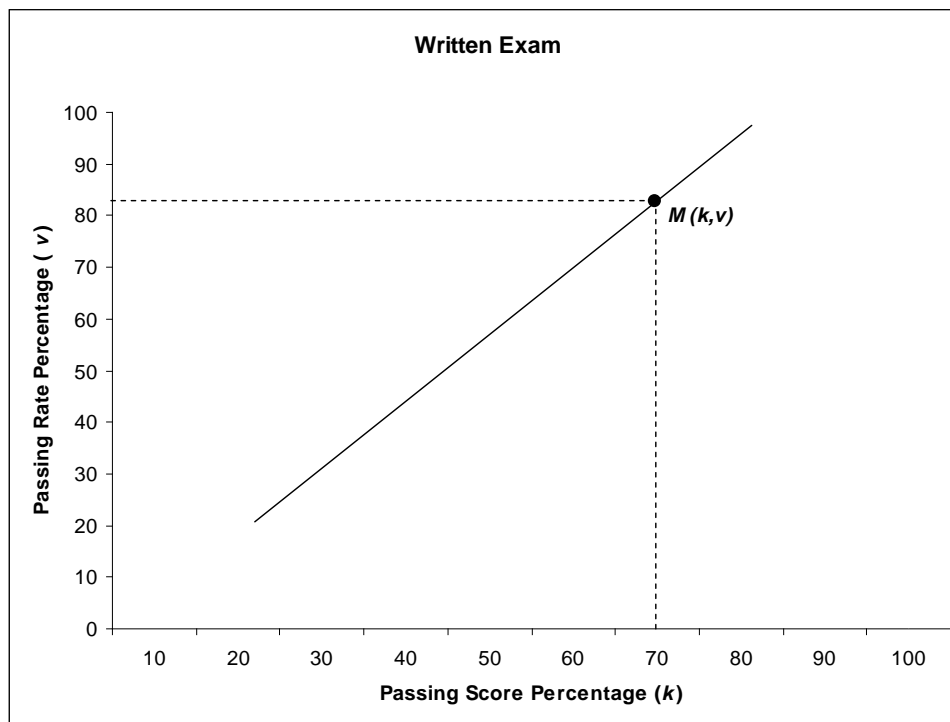


Figure 2. Line with slope s_v/s_k has been drawn through point M.

Table 3 summarizes the empirical passing rates for all three tests. The column labeled 'Interval' contains values representing a theoretical passing score. Passing rate columns contain the rate of participants who would pass based on the theoretical passing score for each test. For the Written Exam, if the passing rate were set to 20%, approximately 99.1% of the

population used for this study would pass. If the passing score were set to 30%, 98.2% would pass. If the passing score were set to 40%, 90.6% would pass, and so on. Values for the Pre-Operational and Operational Exams were summarized in a similar manner. Intervals were chosen based on the resulting curve that captures

all participants' data and provides a smooth curve with which to perform the adjustment.

In Figure 3, the percentage of examinees who were successful (the passing rate y) was plotted as a decreasing function of the theoretical passing test scores (the passing score x). A basic function in Excel was used to plot the curve of passing rate values against the

interval of theoretical passing scores (see Table 3 for values associated with each exam). The number (N) for each test instrument varied based on successful completion of each stage of testing by the participants. The resulting curve intersected the line that was plotted previously and was also used for the next step in the procedure.

Table 3
Summary of Empirical Passing Rates per Interval Based on Passing Score

Written Passing Rate	Interval	Pre- Operational Passing Rate	Interval	Operational Passing Rate	Interval
99.11	20	100	10	4.65	0
98.23	30	99.72	15	15.50	1
90.56	40	99.72	20	30.23	2
77.58	50	99.45	25	40.31	3
58.11	60	98.10	30	49.61	4
32.15	70	96.20	35	60.47	5
8.84	80	95.93	40	68.99	6
0	90	93.22	45	74.42	7
0	100	91.59	50	78.29	8
		86.99	55	86.05	9
		84.55	60	88.37	10
		76.96	65	89.92	11
		72.08	70	89.92	12
		66.66	75	90.70	13
		58.80	80	93.02	14
		48.50	85	93.80	15
		34.41	90	93.80	16
		21.68	95	93.80	17
		10.02	100	93.80	18
				95.35	19
				99.22	20

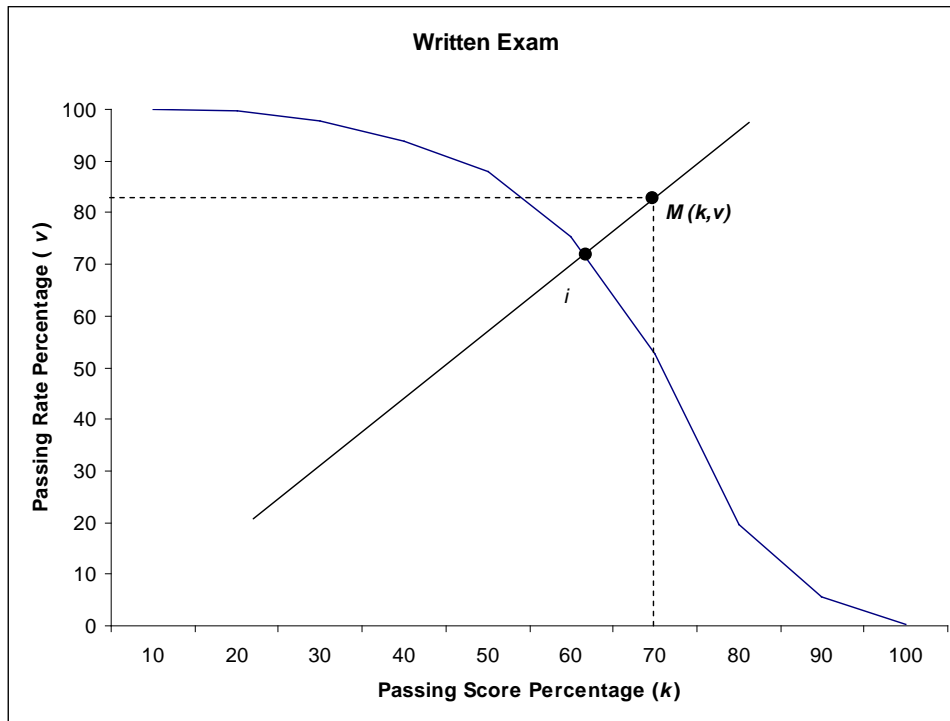


Figure 3. Examinee success plotted as a function of the passing score.

In Figure 4, the point (*i*) where the sloped line intersects the curve is the area where the adjustment occurs. This new point results with a

recommended adjustment to the passing score with the corresponding passing rate based on the new passing score.

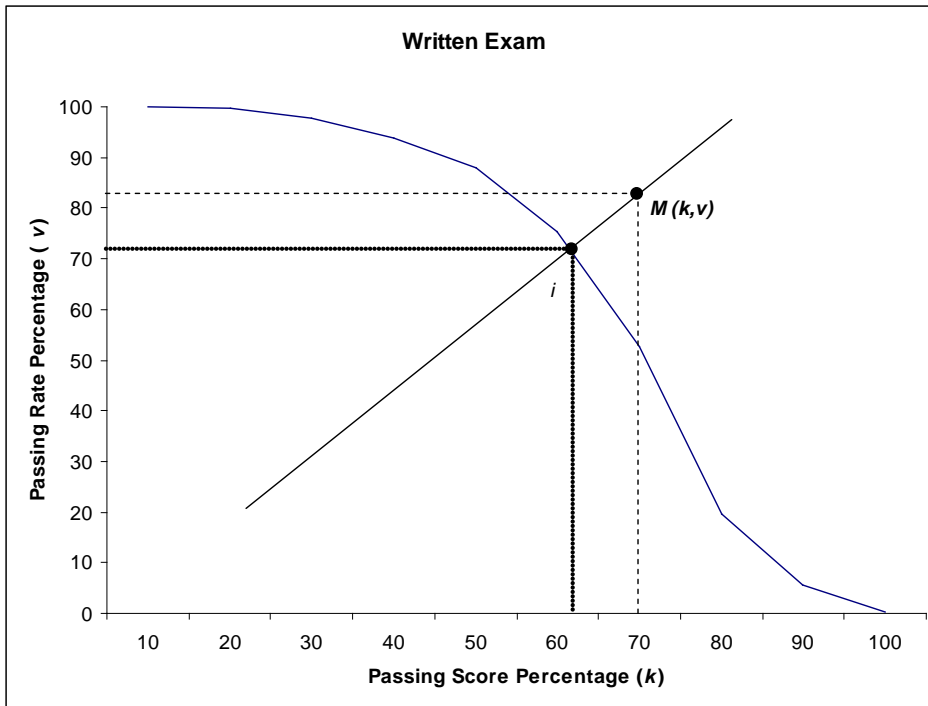


Figure 4. Proposed passing score and passing rate for the Written Exam.

Figure 5 and Figure 6 show the result of the Beuk procedure for the other two exams. Values (panel and empirical) from each exam were used to perform the procedure as with the Written Exam. As can be seen in the graphs, location of

the points, axes, slope of the line, and shape of the empirical data curve are respective to each test resulting in unique outcomes.

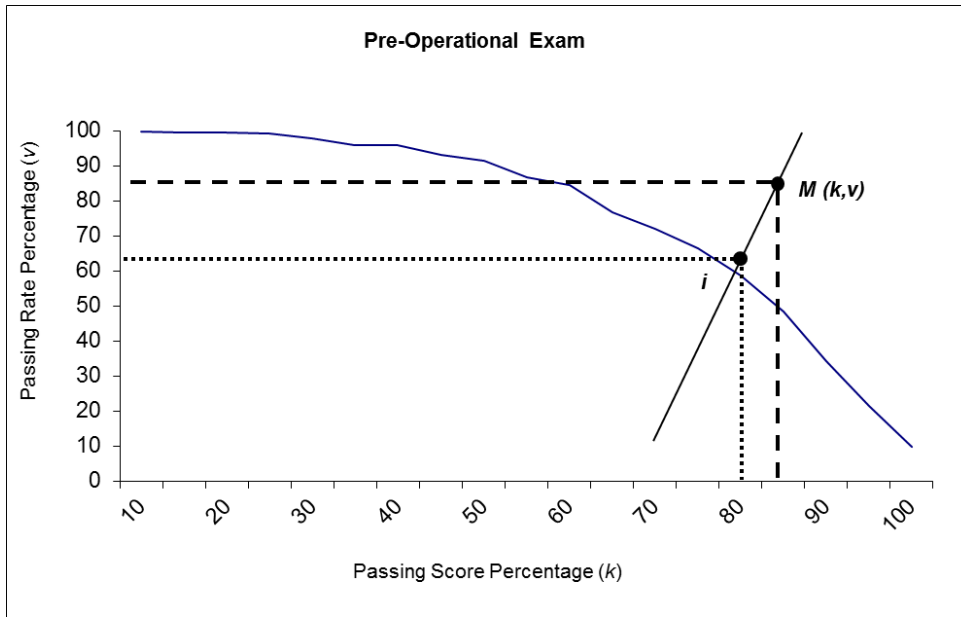


Figure 5. Summary of the Beuk procedure for the Pre-Operational Exam.

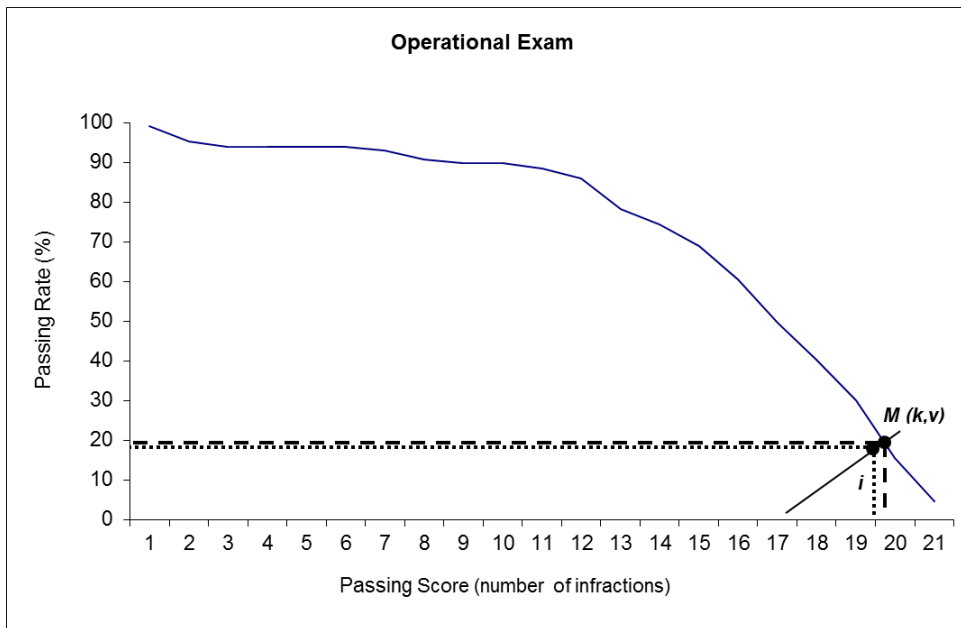


Figure 6. Summary of the Beuk procedure for the Operational Exam.

Table 4 summarizes the suggested changes for all three tests in the Gearing Up for Safety

curriculum. Values from before and after the procedure are included.

Table 4

Summary of Passing Scores and Passing Rates Representing both Original and Recommended Changes from the Adjustment Procedure.

	Written Exam		Pre-Operational Exam		Operational Exam	
	Original	Recommended Change	Original	Recommended Change	Original	Recommended Change
Passing Score	70%	63%	85%	81%	<20 infractions	<19 infractions
Passing Rate	84%	71%	90%	62%	82%	81%

Panel Discussion

The conclusions of the Beuk procedure were brought back before the subject matter experts for review. As with other stages of development for the curriculum, the panel was given the directive to consider adjusting the passing standards for the exams based upon the findings of this study. There were three choices for the panel regarding each of the three test instruments: (a) accept changes as suggested by the procedure, (b) reject changes suggested by the procedure, or (c) accept a change other than one suggested by the procedure. The panel concluded that both the Written and Pre-Operational Exams should retain their original passing scores and voted to reject the suggested change, which would lower the acceptable passing scores. This was due to the consensus that as thresholds in the curriculum, the standard should be higher to more accurately identify participants who need more preparation or are not ready to progress through the remainder of the program. The adjustment procedure suggested that the Operational Exam's score basically remain the same. The panel, however, concluded that the passing score should be adjusted from 'less than 20 infractions' to 'less than 15 infractions' on the premise that participants in the final stage of testing should be able to meet a stricter standard. This decision was also based on reviewing empirical data and discussing participants' performance during the Operational Exam. Concerning the passing scores for the Gearing Up for Safety tests, the panel made a final decision of a 70% passing score for the Written Exam, an 85% passing score for the Pre-Operational Exam, and less

than 15 infractions allowed for the Operational Exam.

Discussion

The Gearing Up for Safety curriculum was designed to evaluate the readiness of youth to operate certain types of agricultural equipment as prescribed by the AgHOs. Participants' progression through the testing process is determined by successfully achieving the established passing scores for each of the three tests administered at the conclusion of each stage. The purpose of this study was to examine the appropriateness of those passing scores utilizing a compromise method (Beuk, 1984). While not commonly found in research to date, re-examining passing standards can provide non-arbitrary information with which to strengthen a curriculum's passing standards and provide more validity evidence for the set criteria. More importantly, it serves as an opportunity for the stakeholders to review the testing program and reflect on the passing rate of the candidates completing the program.

Some important points should be made about establishing and adjusting passing standards. First, passing scores should be viewed as 'estimates' since abilities cannot be objectively measured, there is variability of the construct, and panel make-up can produce variability in establishing the score. Thus a 'true' or 'exact' passing score does not exist and periodic review can result in minor changes (O'Neill et al., 2005; Wendt & Kenny, 2007). Second, using only one source of data is not sufficient for score establishment or adjustment

as noted in the Standards for Educational and Psychological Testing and other research studies (AERA, APA, & NCME, 1999; Cizek, Bunch, & Koons, 2004). Adjustment procedures should involve a balance of expert opinion and empirical data (Beuk, 1984; Mills & Melican, 1988; AERA, APA, & NCME, 1999). When combining more than one source of information and examining a curriculum with a holistic view, developers can come to better conclusions.

Although the group of experts rejected the recommendations for adjusting passing scores (which would have generally reduced the needed achievement level), there is tremendous value in the procedure. Through more informed evaluation of established scores, the panel and researchers were able to review the empirical performance of the three test instruments as a curriculum package for the first time since the Gearing Up for Safety curriculum began over a decade ago. Passing standards, established from previous research, were also strengthened through research and re-evaluation (French, Field et al., 2006, 2007).

The compromise method (Beuk, 1984) provided a useful comparison of research established passing scores for the Gearing Up for Safety exams and empirical data gathered from administering the test to the target population for this study. In fact, it was an evaluation of a theoretical or even ideal standard to that which occurs in practice with the opportunity to re-adjust that standard. The procedure also confirmed work provided by subject matter experts and decisions made about levels of performance. Further testing should be performed using youth in other areas of the United States and at younger ages to be more representative of a national program. Upon

obtaining those results, the compromise procedure should be performed to examine the appropriateness of the scores with a more geographically representative population. While the review did not result in major changes, comparing the original information with the results of this study permitted more informed decisions, resulting in the establishment of more accurate scoring standards. In fact, if the first standard setting process worked well, it could be argued that not making an adjustment through this adjustment study affirms the quality of the first process. This type of evidence adds to the validity argument for the inferences about examinees that are based on the scores from these three exams.

Finally, it should be noted that there has been a general shift, during the past decade, in the social contract that exists between society and the agricultural community. In the past, there has been a much higher tolerance of the practice of children and youth being engaged in certain types of agricultural work. Efforts to promote child safety on farms had focused on discouraging parents from assigning children to certain hazardous tasks. The change in the tolerance of children working in agriculture may have influenced the reluctance of the subject matter experts to lower testing standards for achieving certification for employment under the provisions of the AgHOs. As the resistance to youth being engaged in hazardous agricultural work practices increases due to the relatively higher injury rate for youth in the past, there has been a push towards higher standards as was noted by the consensus of the subject matter experts in this study.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- Beuk, C. H. (1984). A method for reaching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement*, 21(2), 147–152.

- Bullock, S. R. (2006). *Evaluating the effectiveness of a visually-based farm tractor and machinery safety curriculum compared to a text-based curriculum* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.
- Cizek, G. J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement, 30*(2), 93–106.
- Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice, 23*(4), 31–50.
- De Gruijter, D. (1985). Compromise models for establishing examination standards. *Journal of Educational Measurement, 22*(4), 263–269.
- French, B. F., Breidenbach, D. H., Field, W. E., & Tormoehlen, R. L. (2007). The psychometric properties of the Agricultural Hazardous Occupations Order certification training program written examinations. *Journal of Agricultural Education, 48*(4), 11–19. doi: [10.5032/jae.2007.04011](https://doi.org/10.5032/jae.2007.04011)
- French, B. F., Field, W. E., & Tormoehlen, R. L. (2006). Performance standards for the Agricultural Hazardous Occupations Order (AgHOs) certification training program exam. Proceedings from NIFS 2006: *National Institute for Farm Safety Conference Paper No. 06–04*. Sheboygan, WI.
- French, B. F., Field, W. E., & Tormoehlen, R. L. (2007). Proposed performance standards for the Agricultural Hazardous Occupations Order certification training program. *Journal of Agricultural Safety and Health, 13*(3), 285–293.
- Geisinger, K. F. (1991). Using standard-setting data to establish cutoff scores. *Educational Measurement: Issues and Practice, 10*(2), 17–22.
- Kingman, D. M., Yoder, A. M., Hodge, N. S., Ortega, R., & Field, W. E. (2005). Utilizing expert panels in agricultural safety and health research. *Journal of Agricultural Safety and Health, 11*(1), 61–74.
- Mills, C. N., & Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. *Applied Measurement in Education, 1*(3), 261–275.
- O'Neill, T. R., Marks, C. M., & Reynolds, M. (2005). Re-evaluating the NCLEX–RN passing standard. *Journal of Nursing Measurement, 13*(2), 147–165.
- Ortega, R. R. (2003). Analysis and evaluation of the effectiveness of the 4–H CAI/Multimedia Farm Tractor and Machinery Safety certification program (Unpublished master's thesis). Purdue University, West Lafayette, IN.
- United States Department of Labor. (1971). *Code of Federal Regulations Title 29 Chapter V Part 570 Subpart E–1: Occupations in agriculture particularly hazardous for the employment of children below the age of 16*. Retrieved from <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=ea166331fde488bb005379e601a83cf6&rgn=div5&view=text&node=29:3.1.1.1.30&idno=29#29:3.1.1.1.30.6>.
- U.S. Department of Labor. (2007). *Child labor requirements in Agricultural Occupations under the Fair Labor Standards Act (Child Labor Bulletin 102)*. (Employment Standards Administration Wage

and Hour Division Publication 1295). Retrieved from
<http://www.dol.gov/whd/regs/compliance/childlabor102.htm>.

Wendt, A., & Kenny, L. (2007). Setting the passing standards for the National Council Licensure Examinations for Registered Nurses. *Nurse Educator*, 32(3), 104–108.

WILLIAM BRIAN HOOVER is an Assistant Professor of Agricultural Systems Technology in the Department of Agricultural Science at Murray State University, 213 South Oakley Applied Science, Murray, KY 42071, brian.hoover@murraystate.edu

BRIAN F. FRENCH is an Associate Professor of Educational Leadership and Counseling Technology at Washington State University, 362 Cleveland Hall, Pullman, WA, frenchb@wsu.edu

WILLIAM E. FIELD is a Professor of Agricultural & Biological Engineering at Purdue University, 225 South University Street, West Lafayette, IN 47907, field@purdue.edu

ROGER L. TORMOEHLEN is a Professor and Head of the Youth Development and Agricultural Education Department at Purdue University, 615 West State Street, West Lafayette, IN 47907, torm@purdue.edu