Journal of Agricultural Education, 66(4), Article 33
https://doi.org/10.5032/jae.v66i4.3277

Reliability Analysis of Test Questions to Support a National Tractor and Machinery Safety Certification Program

Jill F. Kilanowski[1]
S. Dee Jepsen[2]
Rebecca Berenbon[3]

Abstract

The purpose of the study was to evaluate the reliability of a newly developed test bank of exam questions to support the U.S. Department of Labor tractor and machinery safety certificate's written exam. This test bank was developed by a panel of eight subject matter experts, from different U.S. geographic regions, in face-to-face meetings for item writing, to assess subject knowledge in middle and high school students aged 13-to-18 years old. The guiding theoretical frameworks were Webb's Depth of Knowledge and Bloom's Taxonomy. Via multiple recruitment efforts, high school agricultural educators and community 4-H advisors who taught youth certification programs were recruited to serve as classroom facilitators and replace their current knowledge exam with the study's WebXam online exam. Students (n = 96) from three states and six classrooms participated in the study, representing four different program formats. Results showed strong reliability (a = 0.93) for the exam, with no significant difference in the likelihood of passing based on age or sex. Seven themes were included in the 70-item test bank, where all questions were answered correctly by at least 30% of respondents, indicating there were no questions too difficult for the test-taker population. The results will serve as a foundation for establishing an online testing platform for educators who offer tractor and machinery safety trainings across the U.S., and in-turn may also create program consistency for the DOL written exam with a national test bank of valid and reliable questions.

Introduction

The U.S. Department of Labor (DOL) regulations specify that youth younger than 16 years cannot operate a tractor for hire, only legally operate such equipment on their own family farm (US DOL, 2024). However, if the youth are 14-and-15 years old and successfully complete a tractor and machinery safety course where both written knowledge and hands-on driving are evaluated, they may earn a certificate that enables them to be employed off the family farm (US DOL, 2016a). This legislative exemption, provided in the Agricultural Hazardous Occupations Order (AgHOs), is commonly known as Tractor and Machinery Safety Certification and has been in existence since 1968. The training program plays a crucial role in providing technical education and safety training for young workers in agricultural settings.

The training programs occur in classroom and community-based settings, where the legislation is specific that the only two recognized agencies are the U.S. Office of Education Vocational Agriculture Training Programs and the 4-H Federal Extension Service Training Programs (US DOL, 2016b). However, it is permissible for additional resource persons to assist in the training, including farm machinery representatives, agricultural organizations, or key volunteers knowledgeable of equipment operation. Regardless of which course the youth enroll in to receive their training, having a designated course for safe

[1] Jill F. Kilanowski, PhD, RN, CPNP, FAAN, Research Associate, Dept. of Food, Agricultural & Biological Engineering, College of Food, Agricultural, and Environmental Sciences, 590 Woody Hayes Drive, Columbus, OH 43210, kilanowski.2@osu.edu
[2] S. Dee Jepsen, PhD, Professor, Dept. of Food, Agricultural & Biological Engineering, 590 Woody Hayes Drive, Columbus, OH 43210, jepsen.4@osu.edu
[3] Rebecca Berenbon, PhD, Associate Psychometrician, Center on Education and Training for Employment (CETE) Ohio State University, 1900 Kenny Road, Columbus, OH 43210, berenbon.1@osu.edu

tractor and machinery operation offers teen workers far more than just equipment knowledge. The training builds foundational career skills and sets young workers up for success in the agricultural industry.

Since 2001, funding was provided to land grant institutions to develop and enhance youth agricultural safety training programs from the U.S. Department of Agriculture (USDA) National Institute for Food and Agriculture (NIFA). Several national curricula are available to teach tractor and machinery safety programs, each with their own core competencies and student assessment strategies. Notable programs include the Hazardous Occupations Safety Training in Agriculture (HOSTA) (Murphy, 2019) which was a precursor to the National Safe Tractor and Machinery Operation Program (NSTMOP) (Harshman et al., 2017). This program is now available in a web-based format through AgSafety4U (Yoder, 2020). Another national program is Gearing Up for Safety: Production Agriculture Safety Training (Tormoehlen et al., 2003; Purdue University, n.d.). These curricula were systematically developed to encompass essential content areas, instructional methodologies, and evaluation of metrics, which are foundational components of any effective training course. Empirical research of the tractor and machinery safety training programs includes the definition of core content areas (Ortega et al., 2003; Garvey et al., 2008; Mathew et al., 2011; Snyder et al., 2013; Mann & Jepsen 2017), validation and reliability of student assessments (French et al., 2007; Mann et al., 2016) and determination of minimum passing scores (French et al., 2007; Hoover et al., 2012). Collectively, these studies have significantly enhanced the quality and consistency of instructional courses designed to ensure students are adequately prepared to operate hazardous agricultural equipment safely. Such scholarly efforts were not part of – or were not adequately documented – when legislation was enacted in 1968.

In the original AgHOs, testing criteria was vague and the legislation did not elaborate how students are to show evidence of their core, foundational, and basic safety knowledge, except to say they will pass a written test, followed by a practical test which includes driving (US DOL, 2024). In academia, there is a process to construct a quality written test that includes consideration of the objectives, depth of content coverage, planning a blueprint to guide the design of questions, and presenting questions that span a variety of styles. When a cognitive perspective is considered in test development, the different ways students learn, process and apply knowledge ideally directs questions to be fair, clear, and determines the best item form and response categories (Finn et al., 2014). To write a quality test, a common test writing approach is to utilize subject matter experts (SMEs) to determine the perceived importance of the subject topics, and through consensus, establish content validity of subject matter (Haladyna, 2002; Kingman et al., 2005; Mathew et al., 2011). The SMEs are recognized by their peers to have background and expertise in a specific area and therefore qualified to serve in the research study role for establishing exam items (Meyer & Booker, 2001). Using a consensus approach to exam writing enhances alignment to standardized core competencies and reduces teacher bias for instructional priorities and content selection, which are common elements found in teacher made tests (Wellberg, 2023). Wellberg also reported that current research is deficient in the statistical measurements of quality in teacher made exams.

In a previous study (Mann, 2017), eight subject matter experts with experience in tractor and machinery safety were recruited to create a universal assessment compatible with the minimum core requirements across all national curricula. These SMEs were grounded in agricultural mechanics and geographically dispersed in the U.S. They had access to test questions from NSTMOP, Gearing Up, and other teacher-derived curriculum. Through a systematic analysis, these questions were reviewed and modified to create a new test bank of questions for national use. This process confirmed the content validity of the exam; however, the reliability of the assessment has yet to be established.

Reliability (internal consistency) is measured by Cronbach's alpha (α) and is the degree to which an exam is dependable, consistent, and stable in measuring what it is intended to measure (Nunnally & Bernstein, 1994). In other words, if a person takes the exam again, with a test of high internal consistency, their score should be similar. An exam without reliability is no better than randomly assigning grades to

students. The standard error of measurement (SEM) is a statistic that estimates the variability in a candidate's observed score from a true score (which is defined as the average obtained by the individual over hypothetical repeated testing with the same test form) (Nunnally & Bernstein, 1994). Standard error of measurement accounts for both reliability and score variability, with lower values indicating more precise measurement (defined as a narrower interval around observed scores).

The potential impact of our research is that the written exam for the DOL Tractor and Machinery Safety Certification will be nationally standardized for consistency. Testing students on core content helps verify that students in rural, urban, or suburban schools are learning the same essential knowledge and skills, regardless of the curricula they studied or teaching styles they encountered in the classroom. In addition, the workload of instructors will be eased with them knowing a valid and reliable exam is available for their classroom adoption, ensuring their students are prepared in the same manner as other students entering the workforce.

Conceptual Frameworks

Our study was informed by two theoretical frameworks, namely Webb's Depth of Knowledge (2002) and Bloom's Taxonomy (1956). These two frameworks were chosen as they had application to the construction of the exam, the depth of knowledge needed to be learned by the student, and the most appropriate manner to evaluate the quality of the knowledge assessment tool. Using these models in tandem, item writing of the exam questions appropriately matches the cognate level, and the cognitive complexity needed to successfully answer item questions. Cognate knowledge is important in testing as the cognate level of a test item should not be higher than the task and guiding statement (Jensen et al., 2020). In our written exam we used Webb's Depth of Knowledge Level 1 and Bloom's Cognate Level 1.

The Webb' Depth of Knowledge framework (see Table 1) has categorized student intellectual and academic learning into four levels of cognitive complexity (Webb, 2002; Webb et al., 2005). Instead of focusing on students doing tasks, it focuses on how students must learn to complete a task. It represents a pedagogical standard that supports academic quality and rigor. Testing within this model uses Depth of Knowledge Level 1 (i.e. recall questions) to test for Cognate Level 1, and Depth of Knowledge Levels 2 – 4 for higher cognitive tasks.

Table 1

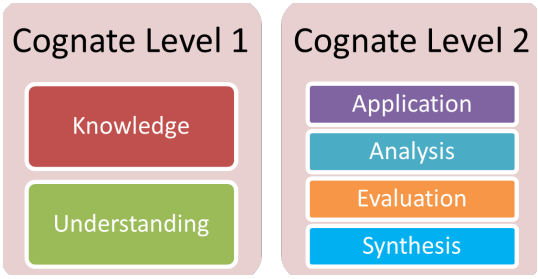
Modified Webb's (2002) Depth of Knowledge Levels

Depth of Knowledge Level (DOK)		Title of Level and Cognitive Demands
1	Describe	Recall and reproduction
2	Explain	Skills and concepts
3	Interpret	Short-term strategic thinking
4		Extended thinking

Bloom's taxonomy (Figure 1) indicates the level of expertise needed for the student participant to achieve the learning outcome in Cognate Level 1 are associated with knowledge and understanding, while Cognate Level 2 is associated with application, analysis, evaluation and synthesis (CETE, 2016). In both the Webb and Bloom frameworks, the classroom educator aligns their learning goals and considers outcomes for success and the appropriate manner to measure that success.

Figure 1

*Modified Bloom's Taxonomy Based on Cognate Levels of Learning**



*With permission from Ohio State University Center on Education and Training for Employment. (2016). *A subject-matter expert's guide. Creating quality multiple-choice items*.

Purpose and Research Objectives

The purpose of this study was to establish item reliability scores for questions within a tractor and machinery safety test bank, from students enrolled in a DOL safety course. An initial reliability pilot assessment conducted with 15 Ohio high school agriculture teachers yielded Cronbach's alpha of 0.892 (Mann, 2017). The research objective was to have tractor safety course instructors in different geographic locations, as well as from different curriculum formats, administer the newly created online test to their students to measure their knowledge and determine their passing rate after completing a tractor and machinery safety course. A second objective was to establish reliability scores and item difficulty analysis for the exam. Ultimately, the goal of our study was to establish a reliable test instrument to assess students' knowledge of safe tractor and machinery concepts and satisfy the federal training process, in natural learning environments of high school classrooms, Extension programs, and independent study settings. This study received USDA-NIFA funding through two Safety in Agricultural for Youth (SAY) grants in 2016 and 2021.

Methods

The research design was quantitative analysis. The sample population was educators of middle and high school students who teach tractor and machinery safety. Initial recruitment targeted educators listed in a public database and contact list used in the primary Mann (2017) study. Using a modified Dillman (2007) method, electronic invitations were sent to agricultural educators in both school and community club settings, inviting them to serve as test facilitators. In this role, educators replaced their standard final exam with a 70-question test bank developed for the study. Each educator received one initial recruitment email followed by two reminder messages.

Recruitment began weeks before the onset of the COVID-19 pandemic, which led to widespread school closures and a rapid shift to online learning. It was assumed that participation in the study became a lower priority during this transition. Despite these recruitment challenges, one dedicated educator continued with the study and successfully administered the exam to 29 students in a virtual classroom setting. As educational institutions gradually returned to in-person instruction, the research team resumed and expanded recruitment efforts. Institutional Review Board (IRB) modifications enabled recruitment through professional organizations and events where participants attended, i.e. agricultural education conferences, teacher workshops, agricultural trade shows, Extension/4-H programs, and National FFA

Convention. Informational booths at these events helped attract additional instructors to receive our invitation to facilitate the test.

Upon consenting to participate, educators were provided with written tutorials on using the WebXam platform, along with online support from both the OSU CETE and the research team. Over a three-year recruitment period, a total of 96 students – both assenting and non-assenting – completed the exam.

The procedures of the study had educators use their choice of curriculum to teach the tractor and machinery safety program content. The courses utilized for instruction included HOSTA, NSTMOP, Gearing Up, AgSafety4U, and the Golden Triangle Curriculum Cooperative (GTCC, n.d.) which is a Montana-based curriculum utilized in rural and private school districts. The test bank included 70 multiple choice questions on topics of: earthmoving machinery operations, front-end loaders, general machinery operations, power take-off (PTO) implement operations, tractor operations, roll-over protection (ROPS), along with miscellaneous transportation, skid steers and risk communication. All test items were formatted as multiple-choice questions with a variety of designs; some questions were scenarios, some showed images, and some recalled facts and instructional information.

Teachers were responsible for determining the test administration date and time. The exam duration was allocated for one hour, and student accommodations were honored. The research team was responsible for setting up each classroom in WebXam while the educators provided the students' name, email address, age at testing, sex and grade. As requested by the university IRB, school principals/program administrators provided letters to waive parental consent for participation in the study. The waiver of parental consent did not adversely affect the rights of the participants because students complete subject content quizzes and examinations as standard educational practice, and to ask for parental consent would be a deviation of current academic practice. Similarly, students do not sign assent to complete subject content quizzes and/or examinations. Nonetheless, students were required to provide an electronic assent to have their exam results included in the study by responding with a yes for study participation to the first question in WebXam. All students, regardless of their assent to participate in the research study, were awarded a digital badge via Badgr (Instructure, 2025) for Tractor and Machinery Safety-Written exam when they achieved a 70% or higher score of material knowledge accuracy. This written exam achievement also served as a step towards the achievement of a DOL national certificate. In all cases, students and their teachers were informed of achievement scores in the exam.

This study worked in collaboration with the logistical support of the Center on Education and Training for Employment (CETE) located at the OSU Columbus campus. The CETE staff provides translational research guidance for curriculum assessment, as well as assists in the development and verification of credentialing examinations using evidence-based interventions for policy development and adaptation of educational practice (OSU CETE, 2025). WebXam (CETE, n.d.) is developed and owned by OSU's CETE and is a secure web-based assessment delivery system. It is used in educational and employment settings in local, regional, and national arenas. WebXam has the capability to have instant scoring and can report results on an individual and aggregate level. WebXam has university levels of research integrity with a secure network utilizing two-step user and computer verification. All members of the research team and CETE staff have Collaborative Institutional Training Initiative (CITI) training to maintain the protection of human subjects (CITI, 2025). Furthermore, CETE enabled the logistical awarding of credentialed digital badges though Badgr (Instructure, 2025).

The CETE staff accepted leadership in the item analysis of WebXam questions using the computer software Lertap, (a program used to process and analyze data collected from quizzes, exams, tests, and surveys) and generated summary reports. Lertap is an acronym for the Laboratory of Educational Research Test Analysis Package (Assessment Systems Corporation, 1979). Student descriptive data, item difficulty,

item discrimination, and regression analyses of demographic data to exam scores were conducted. The study and its addendums were approved by the university IRB 2023B0141.

Findings

In total, 96 students in grades 8-12 from three different states completed the tractor and machinery safety exam in WebXam with 89 (93%) assenting to participate in the research study. Of those 89 students, 63 (70.79%) achieved a score of at least 70%. The average or mean score for this exam was 50.83 points (72.62%) and the median or middle score was 55 points (78.6%). There were no missing data, and all students received a numerical score. The minimum observed score was 16 points, and the maximum observed score was 70 points out of a possible 70. This suggests a "ceiling effect" in the score distribution, meaning the test may not have measured the full range of abilities in the test-taker population. Scores of 100% were achieved in two settings, one in a 4-H setting with an educator who taught the NSTMOP tractor and machinery safety course for over 30 years, and one student who was home-schooled and enrolled in the online AgSafety4U curriculum. Of the seven students who did not assent, one student earned a passing score, while the others did not pass; no further analyses were conducted on these seven students. Table 2 shows how the various research sites performed with the number of students enrolled, assented, and levels of success. The demographic descriptions (age, grade and sex with scores) of the assenting students are in Table 3.

Table 2

Student enrollment, assent, and test success with state and program description

Location	Total students	Not assented	Assented	Passed Assented	Passing rate of assented	Curriculum
Montana HS*	29	5	24	6	25%	Golden Triangle Curriculum Cooperative
Ohio HS	18	1^	17	11	65%	Hazardous Occupations Safety Training (HOSTA)
Ohio 4-H	10	0	10	10	100%	National safe tractor & machinery operation program (NSTMOP)
KY self-study	1	0	1	1	100%	AgSafety4U online self-study
Ohio – 8 & 9 grades	6	0	6	5^	83%	National safe tractor & machinery operation program (NSTMOP) + teacher's own resources
Ohio HS	32	1#	31	30	97%	Gearing Up for Safety
Summary	96	7	89	63	⊕/⊕⊕	⊕ 71% of ALL assented students passed ⊕⊕ Removing Montana (Covid & GTCC) 88% of assented students passed.

*Administered during COVID online classroom – see additional information for this classroom
^ Unsuccessful student had academic accommodations
# Student who did not assent, did pass

Table 3

Demographics of test participants with score, Standard Deviation (SD), and pass rate percentage

Variable	*n*	%	Score *M*	Score *SD*	Students' Pass Rate (%) *
Age (years)					
13	7	7.87	44.43	14.51	57.14
14	8	8.99	53.38	12.40	62.50
15	21	23.60	52.71	8.00	80.95
16	31	34.83	52.19	12.84	74.19
17	15	16.85	48.00	15.56	66.67
18	7	7.87	48.71	12.08	42.86
Grade					
8th	12	13.48	46.50	13.98	50.00
9th	8	8.99	52.75	10.22	75.00
10th	35	39.33	53.06	10.94	80.00
11th	18	20.22	50.33	15.75	72.22
12th	16	17.98	48.81	10.91	56.25
Sex					
Male	61	68.54	51.54	13.56	72.13
Female	28	31.46	49.29	9.29	64.29
Overall	89	100	50.83	12.29	69.66

*Note. Students' pass rate indicates the percentage of students who scored 70% or higher.

The pass rate for the Montana students who took the exam in the early months of the COVID pandemic, while forced into a complete online learning platform of the GTCC curriculum, was 25.00% (meaning 6 out of 24 students passed). Their average percent score was 54.52% and noteworthy; their teacher informed the study team that students had no consequences for not completing classroom assignments or neglecting online class participation during the pandemic.

The Cronbach's alpha (α) of the exam has an estimated reliability of .93, which is excellent according to common benchmarks (Taber, 2018). Only two test items have a higher Cronbach's alpha-if-item-deleted value (.933) than the test reliability (.932), meaning these items only slightly detract from the test's overall reliability. The standard error of measurement (SEM) was 3.22.

Classical test theory was used to evaluate test item difficulty, that is the proportion of test-takers who answered a question or item correctly. A difficulty of .90 means the item was answered correctly by 90% of candidates and thus is easier, whereas a difficulty of .10 means the item was answered correctly by 10% of the candidates and thus is harder. The ideal test item difficulty on a criterion-referenced test is .30-.85 to maximize variation in test scores, thus improving reliability (Crocker & Algina, 1986). A small portion of the test items (7, or 10% of all items) were answered correctly by 90% or more of the test-takers. The themes of these questions focused on front-end loaders (3), power take-offs (PTO) (3), and feed grinders (1). All test items were answered correctly by at least 30% of respondents, indicating there were no questions that were too difficult for the test-taker population.

Item discrimination indicates how well the item differentiates between higher and lower scorers, and point-biserial correlation was used in its evaluation. The range is from -1.0 to 1.0 with higher, positive numbers indicating better-performing items. A point-biserial correlation that is negative or low positive (< .15) is a poorly performing item, while above .15 indicates an acceptable item and above .3 indicates a stronger performing item. Five test items fell below .15, each from different themes of hand signals and tractor and machinery operations. See Figure 2 for a scatterplot of item discrimination and difficulty based on theme labels. Table 4 reports themes of questions with number of exam items in each theme, and range of difficulty and discrimination in each question theme. Table 5 shows individual test item statistics.

Figure 2

Scatterplot of Item Discrimination and Difficulty

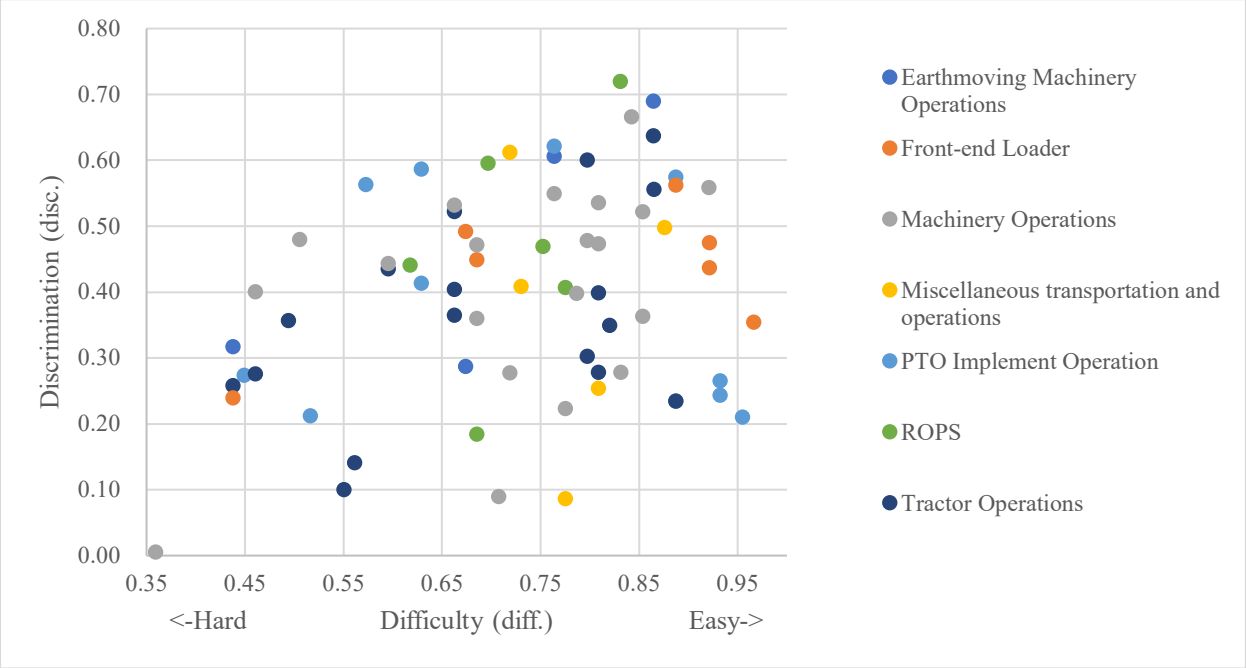


Table 4

Themes of questions, number of questions in theme, range of difficulty and discrimination

Themes	Number of items in theme	Range of difficulty	Range of discrimination
Earthmoving Machinery Operations	4	.44-.87	.29-.69
Front-end Loaders	7	.44-.97	.24-.56
Machinery Operations	20	.36-.92	.01-.67
Miscellaneous transportation and operations	5	.72-.88	.09-.61
PTO Implement Operations	11	.45-.96	.21-.62
Roll-Over Protection (ROPS)	6	.62-.83	.18-.72
Tractor Operations	17	.44-.89	.10-.64

Table 5

Individual test item statistics

Item #	Theme Name	Difficulty	Discrimination
1	Machinery Operations	0.46	0.40
2	Machinery Operations	0.78	0.22
3	Machinery Operations	0.36	0.01
4	Machinery Operations	0.76	0.55
5	Machinery Operations	0.80	0.48
6	PTO Implement Operations	0.63	0.59
7	PTO Implement Operations	0.76	0.62
8	PTO Implement Operations	0.52	0.21
9	PTO Implement Operations	0.45	0.27
10	PTO Implement Operations	0.63	0.41
11	PTO Implement Operations	0.89	0.57
12	Machinery Operations	0.60	0.44
13	Machinery Operations	0.51	0.48
14	Machinery Operations	0.71	0.09
15	Machinery Operations	0.85	0.52
16	Machinery Operations	0.69	0.47
17	Tractor Operations	0.80	0.30
18	Tractor Operations	0.82	0.35
19	Tractor Operations	0.66	0.52
20	Tractor Operations	0.87	0.64
21	Tractor Operations	0.66	0.37
22	Tractor Operations	0.87	0.56
23	Tractor Operations	0.80	0.60
24	Earthmoving Machinery Operations	0.44	0.32
25	Front-end Loaders	0.67	0.49

Item #	Theme Name	Difficulty	Discrimination
26	Earthmoving Machinery Operations	0.87	0.69
27	ROPS	0.75	0.47
28	ROPS	0.83	0.72
29	Earthmoving Machinery Operations	0.67	0.29
30	Tractor Operations	0.46	0.28
31	Tractor Operations	0.55	0.10
32	ROPS	0.62	0.44
33	Machinery Operations	0.81	0.54
34	Machinery Operations	0.72	0.28
35	Front-end Loaders	0.92	0.47
36	Front-end Loaders	0.92	0.44
37	Front-end Loaders	0.97	0.35
38	Miscellaneous transportation and operations: Passenger transportation	0.88	0.50
39	PTO Implement Operations	0.89	0.23
40	PTO Implement Operations	0.96	0.21
41	PTO Implement Operations	0.93	0.27
42	PTO Implement Operations	0.93	0.24
43	Machinery Operations	0.69	0.36
44	Machinery Operations	0.85	0.36
45	Machinery Operations	0.83	0.28
46	Machinery Operations	0.79	0.40
47	Tractor Operations	0.89	0.23
48	Miscellaneous transportation and operations: Risk communication	0.78	0.09
49	Miscellaneous transportation and operations: Risk communication	0.81	0.25
50	Miscellaneous transportation and operations: Risk communication	0.73	0.41
51	Tractor Operations	0.56	0.14
52	Front-end Loaders	0.89	0.56
53	Front-end Loaders	0.69	0.45
54	Tractor Operations	0.49	0.36
55	ROPS	0.70	0.60
56	Tractor Operations	0.81	0.28
57	Tractor Operations	0.60	0.44
58	ROPS	0.78	0.41
59	ROPS	0.69	0.18
60	Tractor Operations	0.44	0.26
61	Miscellaneous transportation and operations: Skid-steer Operations	0.72	0.61
62	Machinery Operations	0.84	0.67
63	Tractor Operations	0.81	0.40

Item #	Theme Name	Difficulty	Discrimination
64	Tractor Operations	0.66	0.40
65	PTO Implement Operations	0.57	0.56
66	Front-end Loaders	0.44	0.24
67	Earthmoving Machinery Operations	0.76	0.61
68	Machinery Operations	0.81	0.47
69	Machinery Operations	0.66	0.53
70	Machinery Operations	0.92	0.56

In addition, two regression analyses sought to investigate differences and patterns in score by age and sex. The first analysis was a linear regression with raw score as the dependent variable. The second analysis was logistic regression, using the students' pass/fail status as the outcome. For both analyses, sex (female) and mean-centered age were included as predictor variables. Grade was excluded to avoid multicollinearity with age, as the two were highly correlated (Spearman's $\rho = .89$, $p < .001$). In the linear regression, the model explained only 0.8% of variance in test scores ($R^2 = .008$). Neither age ($\beta = -.002$, $p = .885$) nor sex ($\beta = -.032$, $p = .427$) were predictive of test scores. This indicates that there were no significant predictive association from age or sex on test score. Results were similar for the logistic regression predicting passing status, showing a pseudo-R^2 value of .013, indicating that age and sex contributed little to the model as explanatory variables. Again, neither age (OR = .91, $p = .600$) nor sex (OR = .688, $p = .444$) were predictive of students' success on the exam. When examining descriptive statistics, we noticed certain age groups had lower or higher test passing rates, but these differences may be due to random error associated with small sample sizes. The regression results suggest that there was no significant difference in the likelihood of passing based on age or sex.

Conclusion, Discussion, Recommendations

This study evaluated the reliability of an existing test bank designed to assess basic knowledge of agricultural tractor and machinery safety. While the test bank had already been established for content validity, reliability scores ensure consistency and stability of the test results over time and across different groups of test-takers. Psychometric analyses demonstrated that the exam is of high quality in its current form, supported by strong reliability ($\alpha = 0.93$; Taber, 2018) and robust item statistics, including measures of difficulty and discrimination (Crocker & Algina, 1986).

Further statistical analysis of item difficulty, discrimination, and distractor effectiveness suggests areas for potential improvement to enhance the quality of the exam. In terms of difficulty, seven questions were identified as potentially too easy, with difficulty indices ($p > 0.90$), indicating the questions should be considered for repair and revision. However, it is also possible that these seven questions reflect safety concepts that are well-integrated into students' behavior, making any question on the topic inherently easy to answer. To maintain the integrity of these items, distractors could be revised to remain incorrect but appear more plausible to test-takers who lack the relevant knowledge. Regarding discrimination, five questions showed weak performance in differentiating between high- and low-scoring students, with point biserial values ranging from 0.01 to 0.14. These items could be considered for replacement or revised to improve their ability to distinguish levels of understanding. Despite these areas for improvement, the majority of items demonstrated strong discrimination and a well-balanced range of difficulty levels. Overall, the test bank exhibits high quality and can be confidently used as a standardized replacement for individual classroom exams in tractor and machinery safety certification.

With regards to achieving a passing score, student performance varied across programs. A consistent pattern emerged with the NSTMOP and Gearing Up for Safety curricula yielding the highest

percentage of students scoring 70% or higher on the exam - successfully meeting the written requirement for a DOL certificate. The average student test score, excluding Montana students, was 79.30%. The pass rate, excluding the Montana classroom, was 86.15% (56 out of 65 students); this pass percentile complements the ideal pass rate of 82% of total test takers determined by Hoover et al. (2012).

It is important to note that the written test is the initial assessment of a student's comprehension of safety and is the first evaluation before they progress to the skill components of the certification. While it is difficult for any exam to thoroughly assess safety knowledge, these results suggest confidence that the test bank effectively measured students' basic knowledge of tractor and machinery safety across content within various national curricula. Additionally, because teachers did not have prior access to the test questions, instruction could not be tailored to specific test items. In our study, teaching to the test where curricula and instruction focus on specific test items or clones (of students) was not possible (Jennings & Bearak, 2014).

In review of Bloom's (1956) framework and Webb's Depth of Knowledge (2002), we note that our written exam asks for recall of facts and procedures which are at Cognate 1/ Level 1. This knowledge is necessary for - and combines with - the next step of DOL testing where students are evaluated on their skills to attach various implements and drive a tractor with a 2-wheeled implement through a prescribed course, Cognate 2/Level 2. Together, the two-part assessment of written and skills tests challenges students into a higher order of thinking, which ultimately leads to a deeper understanding of safety principles.

The generalizability of the results may be limited to the agricultural science educators and community 4-H leaders who choose to offer DOL tractor and machinery safety certificate courses and the limited number of students who enroll in the course. There were limitations affecting the recruitment of instructors into the study with IRB requirements to obtain letters waiving parental consent from principals or administrators of the school or 4-H program; so, while some enthusiastic teachers wanted to facilitate the exam, we experienced situations where the administrators' letters were never received. Within the timeline of our study, new states' legislation restricted where and who has access to student directory information, and this too prevented participation. Finally, some current tractor and machinery safety classroom settings either do not have the availability of computer usage with Internet WebXam access and/or the final content exam is administered via paper and pencil. Furthermore, with a larger sample size, the SEM may have decreased.

Future work in collaboration with CETE will create a website with links to the tractor and machinery safety exam on WebXam so the integrity of the questions will be maintained. Statistically, by using the WebXam software, individual test questions cannot be modified by educators, maintaining the validity and reliability of the exam. Future recommendations may include the development of teacher training in how to use the exam and how to align their curriculum with the exam. Future research may also include exploration of knowledge retention over time and mastery of safety skills. The Ohio State team and the SAY network will be consulted to establish the logistics of the implementation and dissemination including conversations on cost, enabling a standardized nation-wide ready exam for teachers to use for students' enrollment in the DOL tractor and machinery safety certificate program.

This is a report of the research study employed to evaluate the reliability of a valid test bank of questions about tractor and machinery safety. Based on the findings, the highly reliable developed test bank has an appropriate distribution of item difficulties and adequately separated the low performers from the high performers. There is benefit to establish a universal test bank of safety questions to provide program consistency in national DOL certificate programs. A uniform exam also provides educators with a pre-established written exam that measures core safety principles and reduces teacher biased grading, which ultimately supports fairness and evenness to the achievement of DOL certification. In this manner,

instructors do not need to spend time creating and grading their own written assessments, allowing more time for mentoring and evaluating students as they operate equipment during the skill tests.

Agricultural education, with a focus on tractor and machinery operation, plays a key role in preparing students to meet the diverse and complex challenges of modern agriculture. For teachers in traditional agricultural classrooms, as well as urban or inner-city environments using compact tractors, and for students who have machinery operation as part of their Supervised Ag Experiences (SAEs), there is value to have DOL certification. This training creates a pathway for early workforce entry, allowing young workers to gain experience and build responsibility while creating a culture of safety. Having basic foundational knowledge, and tested on real-world competencies, reinforces technical education through experiential learning. This training program can influence their future careers in agriculture or related industries where heavy equipment operation is required.

References

Agricultural Safety and Health eXtension Community of Practice. (2019). *Farm Safety Makes you the Hero. NSTMOP Program Overview.* Retrieved May 12, 2025, from https://ag-safety.extension.org/nstmop-program-overview/

Agricultural Safety and Health eXtension Community of Practice. (2025). *AgSafety4U Certificate Course.* Retrieved May 13, 2025, from https://ag-safety.extension.org/agsafety4u-certificate-course/

Arduini-Van, N. (2020). *Educational Psychology.* Press Books. https://edpsych.pressbooks.sunycreate.cloud/chapter/teacher-made-assessments/

Assessment Systems Corporation. (1979). *Laboratory of Educational Research Test Analysis Package* (Lertap5). (Version 11) [Computer software]. https://lertap5.com/HTMLHelp/HTML/index.html?how_to_get_lertap.htm

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. National Academy of Education. https://naeducation.org/wp-content/uploads/2020/06/Comparability-of-Large-Scale-Educational-Assessments.pdf

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Vol. Handbook I: Cognitive domain. David McKay Company.

Center on Education and Teaching for Employment (CETE). (n.d.). *About WebXam*. Retrieved January 11, 2025, from https://www.webxam.org/Admin/Docs#:~:text=WebXam%20is%20an%20online%20testing%20fulfillment%20of%20graduation%20requirements

Center on Education and Teaching for Employment (CETE). (2016). *A subject-matter expert's guide. Creating quality multiple-choice items*. The Ohio State University, Center on Education and Teaching for Employment.

Center on Education and Teaching for Employment (CETE). (2025). *CETE Homepage*. Retrieved May 23, 2025, from https://cete.osu.edu

Collaborative Institutional Training Initiative (CITI). (2025). *The trusted standard in research, ethics, compliance, and safety training*. Retrieved January 13, 2025, from https://about.citiprogram.org

College of Food, Agricultural, and Environmental Sciences. (2024). *Ohio State University Extension is…*Retrieved January 12, 2025, from https://extension.osu.edu/about/mission-vision-values/osu-extension.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory.* Holt, Rinehart, & Winston.

Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design method-2007 Update with new Internet, visual, and mixed-mode guide*. John Wiley & Sons.

Farm Safety SAY Clearinghouse. (2025). Retrieved March 5, 2025, from https://extension.psu.edu/business-and-operations/farm-safety/say-project.

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C.E., Martin RE, Sheridan, M. A., Gabrieli, C. F., Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science, 25*(3), 736-44. https://doi.org/ 10.1177/0956797613516008.

French, B. F.; Breidenbach, D. H., Field, W. E., Tormoehlen, R. (2007). The psychometric properties of the Agricultural Hazardous Occupations Order Certification Training Program written examinations. *Journal of Agricultural Education, 48*(4), 11-19. https://doi.org/10.5032/jae.2007.04011

Garvey, P. M., Murphy, D. J., Yoder, A. M., & Hilton, J. W. (2008). National Safe Tractor and Machinery Operation Program: Development and content evaluation. *Journal of Agricultural Safety and Health, 14*, 333–349. https://doi.org/10.13031/2013.24567

Gearing Up for Safety Production. (2021). Retrieved March 5, 2025, from https://www.asec.purdue.edu/tractor/index.html

Golden Triangle Curriculum Consortium. *Golden Triangle Curriculum Consortium.* (n.d.). Retrieved May 22, 2025, from https://www.gtccmt.org

Haladyna, T. (2002). *Supporting documentation: Assuring more valid test score interpretations and uses.* Lawrence Erlbaum Associates.

Harshman, W. C., Yoder, A. M., Hilton, J. W., & Murphy, D. J. (Eds.). (Revised 3rd ed.). 2020. *National Safe Tractor and Machinery Operation Program: Student manual*. Penn State Extension, Ohio State Extension, and the National Safety Council.

Hazardous Occupations Safety Training for Agriculture Program (HOSTA). (n.d.). Retrieved Sept 22, 2005, from https://www.nifa.usda.gov/sites/default/files/resources/Hosta_brochure.pdf

Hoover, W. B., French B. F., Field, W. E., Tormoehlen R. L., (2012). Adjusting the passing scores for Gearing Up for Safety: Production agriculture safety training for youth curriculum test Instruments. *Journal of Agricultural Education, 53*(3), 1-13. https://doi.org/10.5032/jae.2012.03001

Instructure. (2025). *Canvas credential, the all-in-one digital badging for higher education*. Retrieved May 23, 2025, from https://www.instructure.com/higher-education/products/canvas/canvas-credentials

Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher, 43*(8), 381-389. https://doi.org/10.3102/0013189X14554449

Jensen, J.L., McDaniel, M. A., Kummer, T. A., Godoy, P. D. D. M., St Clair, B. (2020). Testing effect on high-level cognitive skills. *CBE - Life Science Education, 19*(3), 19:ar39.1-13. https://doi.org/10.1187/cbe.19-10-0193

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*, 61-74. https://doi.org/10.13031/2013.17897

Mann, A. J. (2017). *Identification of learning outcomes and development of assessment methods for agricultural safety and health content in secondary agricultural education classrooms* [Doctoral dissertation, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1492733303282755

Mann, A.J., Field, W.E., Tormoehlen, R., & French, B.F. (2016). Development and validation of written exam items for the Agricultural Hazardous Occupations Orders (AgHOs) Certification Training Program. *Journal of Agricultural Safety and Health, 22*, 121-134. https://doi.org/10.13031/jash.22.11296

Mann, A. J., & Jepsen, S.D. (2017). Using expert panel data to guide youth agricultural safety and health training resources in the US. *Safety, 3*, 1(4). https://doi.org/10.3390/safety3010004

Mann, A. J., & Jepsen, S. D. (2019). Hazardous agricultural tasks completed by youth as part of their Supervised Agricultural Experience (SAE): A descriptive study**.** *Journal of Agricultural Safety and Health, 25*(3), 107–116. https://doi.org/10.13031/jash.12998 [elibrary.asabe.org]

Mathew, S. N., Field, W. E., & French, B. F. (2011). Content validation using an expert panel: Assessment process for assistive technology adopted by farmers with disabilities. *Journal of Agricultural Safety and Health*, *17*, 227–241. https://doi.org/10.13031/2013.38184

Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment: a practical guide*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718485

Murphy, D.M. (2019). *National Safe Tractor and Machinery Operation Program.* Retrieved May 22, 2025, from https://www.academia.edu/?h=39958622

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.

Ohio State University (OSU) Center on Education and Training for Employment (CETE). (2025). *About.* Retrieved April 16, 2025, from https://cete.osu.edu/about/

Ortega, R. R., Tormoehlen, R. L., Field, W. E., Balschweid, M., & Machtmes, K. L. (2003). Determining critical subject matter content for a safety certification program for youth employed in agricultural production. *Journal of Agricultural Education, 44*(4), 67–79. https://doi.org/10.5032/jae.2003.04067

Safety in Agriculture for Youth (SAY) National Clearinghouse. (2025). Retrieved May 10, 2025, from https://ag-safety.extension.org/say-national-clearinghouse/

Pate, M. L., Lawver, R. G., & Sorensen, T. J. (2016). Safety in high school supervised agricultural experiences: Teachers' training and students' injury awareness. *Journal of Agricultural Safety and Health, 2*2(1), 61–74. https://doi.org/10.13031/jash.22.11142

Purdue University. [n.d.] *Gearing Up for Safety: Production Agriculture Safety Training*. Retrieved May 22, 2025, from https://www.asec.purdue.edu/tractor/index.html

Snyder, S., French, B., Field, W., Tormoehlen, R., & Ess, D. (2013). Identification and validation of agricultural hazardous occupations order certification program instructor criteria and competencies. *Journal of Agricultural Education*, *54*(1), 111–123. https://doi.org/10.5032/jae.2013.01111 [files.eric.ed.gov]

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*. *48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Tormoehlen, R., Field, W., Fox, R., Personett, C., Vollmer, W., & Ortega, R. (2003). *Gearing Up for Safety: Production Agricultural Safety Training for Youth*. Purdue University: West Lafayette, IN.

U.S. Department of Labor (DOL). (2024). *Code of Federal Regulations, Title 29, Part 570, Subpart E-1—Occupations in agriculture particularly hazardous for the employment of children below the age of 16 years* (29 CFR § 570 Subpart E-1). U.S. Government Publishing Office. Retrieved May 12, 2025, from https://www.ecfr.gov/current/title-29/part-570/subpart-E-1

U. S. Department of Labor (DOL). (2016a). *Fact Sheet #40: Overview of Youth Employment (Child Labor) Provisions of the Fair Labor Standards Act (FLSA) for Agricultural Occupations,* Revised December 2016, Retrieved May 12, 2025, from https://www.dol.gov/agencies/whd/fact-sheets/40-child-labor-farms

U.S. Department of Labor (DOL). (2016b). Child labor requirements in agricultural occupations under the Fair Labor Standards Act. *Child Labor Bulletin* 102. Publication No. WH-1295. Washington, D.C.: U.S. Department of Labor. Retrieved May 12, 2025, from https://www.dol.gov/sites/dolgov/files/WHD/legacy/files/childlabor102.pdf

Webb, N. (2002). *Depth-of-Knowledge Levels for four content areas*. Madison: Wisconsin Center for Educational Research. Retrieved June 1, 2025, from https://dpi.wi.gov/sites/default/files/imce/assessment/pdf/All%20content%20areas%20%20DOK%20levels.pdf

Webb, N. L., Alt, M., Ely, R., Vesperman, B. (2005, 24 July). *Web alignment tool.* Wisconsin Center of Educational Research. University of Wisconsin-Madison. February 2, 2006. Retrieved June 1, 2025, from https://www.scribd.com/document/280830775/Webb-Et-Al-2005-Web-Alignment-Tool-WAT-Training-Manual-2-1-Draft-091205

Wellberg, S. (2023) Teacher-made tests: why they matter and a framework for analysing mathematics exams, *Assessment in Education: Principles, Policy & Practice, 30 (*1), 53-75, https://doi.org/10.1080/0969594X.2023.2189565

Yoder, A. M. (n.d.). *AgSafety4U 2.0: Agricultural safety and health online training program*. Agricultural Safety and Health eXtension Community of Practice. Retrieved from https://ag-safety.extension.org/agsafety4u-certificate-course/