



MISSISSIPPI STATE
UNIVERSITY™



Using performance-based assessment as an authentic measure of student learning in career and technical education programs

Ashley Priebe Brown
Mississippi State University

LeAnn Gupton Miller
Mississippi State University

The purpose of career and technical education (CTE) is to prepare students for careers. CTE programs are project-oriented and problem-based to provide students with career-ready skills, but current formal assessments for many CTE programs do not authentically measure this type of learning. In Mississippi, performance-based assessments (PBAs) have been developed and piloted to align coursework with assessment, provide an authentic assessment of the skills students acquire through the CTE programs, and obtain an accurate picture of students' career-readiness.

Four CTE programs in Mississippi piloted PBA with their second-year students (program completers). Models of PBA from other states were studied to create a framework for Mississippi's PBA. Lead teachers from around the state partnered with Mississippi State University's Research and Curriculum Unit (RCU) curriculum and assessment specialists to develop the test materials and their corresponding criterion rubrics. Training was provided to those administering the assessment, and the first administration was completed in April 2013. Following the April testing, the pilot participants were surveyed to learn their perceptions of this assessment and to determine a plan for future PBA implementation. Score data were analyzed to determine reliability.

Results indicated that this type of assessment does more closely align with CTE curricula. The majority of the feedback received was regarding process more than content.

As more data are collected, the results can be compared with data from previous assessment instruments to measure the efficacy and validity of PBA. Incorporating suggestions gathered from the survey and other feedback into the assessment will improve the process for future administrations of this type of assessment.

Introduction

With the implementation of the Common Core State Standards, today's discussions of standardized accountability testing tend to focus on the academic realm, specifically language arts and mathematics. Methods for assessing career and technical education (CTE) students are

discussed much less frequently, but these programs are not exempt from accountability measures. In Mississippi, the CTE programs are funded by the Perkins grant. One of the accountability requirements of this grant is the measurement of technical skills attainment, an equity check to ensure that technical skills are being learned across the CTE programs in the state. Until recently, this measure was taken using electronically delivered multiple-choice assessments, which provide an efficient way to test a large number of students. While there is certainly plenty of information taught in CTE courses that can be assessed this way, CTE focuses heavily on the learning and practice of skills that will be used in careers. Many of these skills are difficult to measure using a multiple-choice assessment as they require a more hands-on approach.

The majority of accountability testing has centered on multiple-choice assessments, which are cost-effective to deliver and score on a large scale. While assessments used for accountability are understandably important to maintaining a level of quality in any education system, one of the consequences of assessing this way is that schools feel pressure to perform on these tests. Unfortunately, this can cause authentic instruction to be replaced with teaching to the test, focusing on the practice of multiple-choice items which Darling-Hammond and Adamson (2013) say “encourage[s] recognition of simple right answers rather than production of ideas” (p. 3). Using only multiple choice assessments can be a detriment to the education system as a test can never completely measure all that should be happening in a classroom and should not be used to limit the scope of what is taught in a course. When assessment drives instruction, “it is of paramount importance that tests actually test students on the deeper learning skills that they require in the 21st century” (p. 4). Utilizing performance-based assessment as opposed to multiple-choice assessment “can support more ambitious teaching and learning” (p. 8). A test should reflect the

skills being used in class as closely as possible, and performance-based assessment for CTE students allows the curriculum to align more closely with the form of assessment. The Office of Career and Technical Education and Workforce Development at the Mississippi Department of Education has vocalized a need for its programs to utilize this type of assessment.

Purpose of the Study

This study aims to measure the efficacy of implementing a performance-based assessment system in four Mississippi CTE programs by analyzing data from the process. We were interested specifically in whether there would be a score difference between the traditional Mississippi Career Planning and Assessment System (MS-CPAS2) exams and PBA results. We also wanted to know if the score distribution would be normal. A survey given after the administration included questions about the PBA process from the school level, inquiring about what did and did not work for each CTE center. The survey also requested suggestions for future PBA administrations. The main focus of research into this process dealt with five major questions:

- 1) Was there a score difference between MS-CPAS2 and PBA?
- 2) What was the distribution of the PBA scores?
- 3) What did and did not work about the PBA process?
- 4) How willing were evaluators to participate in future PBA?
- 5) What suggestions do stakeholders have for improving this process?

Literature Review

The current literature about performance assessments highlights both the importance and the challenge of finding ways to include these types of assessments in accountability models. Darling-Hammond and Adamson (2013) note that there has been a recent push to return to

performance-based assessments, which last experienced widespread popularity in the early 1990s and became less common around the time of the passage of NCLB, which “led to the elimination of most state assessments requiring writing, research, and extended problem solving, and their nearly complete replacement with multiple-choice and short-answer tests” (p. 3). The authors name cost as the driving factor behind the decision to assess using tests that are easily administered and scored (p. 8), calling this “penny wise and pound foolish” and “a recipe for low-quality schooling” (p. 9). They go on to propose that spending should shift toward developing more high-quality assessments.

The influence of state testing on instruction is obvious, and Darling-Hammond and Rustique-Forrester (2005) state that since testing “can exert powerful influences on curriculum and instruction” (p. 290-291) when high-stakes decisions are attached to test scores, this influence has the potential to have a positive impact on instructional quality when the assessments are “thoughtfully structured” and when stakeholders seek “to deepen instruction ... rather than merely attaching test-like items to lessons”(p. 293). The authors note that if the purpose of accountability testing is to improve the quality of education for students, testing should “encourage high-quality learning and teaching” (p. 312). They conclude by offering strategies that could allow accountability assessment to have a positive effect on instruction.

Lane (2010) suggested that making a shift from multiple-choice assessments to performance-based assessments is one way to improve and deepen learning. Alignment between classroom activities and assessment measures is important because “[w]hen students are given the opportunity to work on meaningful, real world tasks in instruction, students have demonstrated improved performance on performance assessments” (p. 60). The author suggested ways to design

Brown, Miller

large-scale performance assessments, listing advances in technology, well-written content standards, and evaluator training as ways to achieve the goal of implementing a performance-based assessment across a large system.

Lai, Wei, Hall, and Fulkerson (2012) noted the return of performance-based assessments and added that using these types of assessments for high-stakes purposes requires “innovation in the identification and collection of validity evidence that can support these new and varied purposes and contexts” (p. 4). If a test is going to be used for teacher evaluation, showing how it aligns with the curriculum is important, and if it is going to be used with consequences to students, it is important to know that the students had the opportunity to learn the content included in the test. The authors proposed a framework for establishing validity for the various uses of performance assessment results. Questions for validating test content, examinee response processes, internal test structure, external relationships, consequences of test score use, scoring, and reliability were proposed to ensure that the test is designed and used to measure an intended outcome.

Linn and Baker (1991) cautioned that consequences of performance-based assessments should be taken into consideration. PBAs could “[lead] to the heavy use of practice materials that closely match the format of the assessment” (p. 6). Directness and transparency are also discussed, directness being the level to which a task measures the skill in question, and transparency being the level to which the individuals using the assessment have an “understanding [of] the basis on which performance will be judged” (p. 8). They proposed eight validation criteria for performance assessments: consequences, fairness, transfer and generalizability, cognitive complexity, content

quality, content coverage, meaningfulness, and cost and efficiency. This list can be used, in conjunction with other criteria, to evaluate the quality of a performance assessment.

Methodology

PBA began in 2010 with what was essentially a portfolio assessment. Students completed tasks throughout the school year and uploaded the work to Blackboard ePortfolios. Due to technological issues, scoring scalability issues, and concerns with how much the portfolio work actually represented student learning, this method of performance-based assessment was retired in the spring of 2012. In August of the same year, PBA was revisited, this time using a structure based on a performance assessment that West Virginia was using at the time. Instead of having students complete a portfolio, they were given a two-part assessment. During the first part, students completed a job application using a resume they had created ahead of time. The second portion gave students a set of curriculum-aligned tasks to complete while being observed and scored by their course instructor plus three external evaluators, who questioned the students about their work once the working-time was finished.

Planning for the spring 2013 administration of PBA began in August 2012. Lead teachers from each Career Pathway were recommended by Mississippi Department of Education (MDE) program supervisors, and these teachers partnered with curriculum and assessment staff from the RCU to provide a framework for performance scenarios and rubrics that would be relevant and fair to each Career Pathway. The curriculum was examined to compare the tasks students were completing in the classroom with what they would be asked to do during the PBA, and a test blueprint was utilized to ensure that all course units were assessed fairly. A rubric was developed and aligned to the units tested. The RCU edited these materials and standardized their format for

delivery via Blackboard. The American Psychological Association Code of Fair Testing Practices was used as a guideline to write a testing manual, and in February 2013, test coordinators attended a training session at Mississippi State University to learn how to facilitate administrations of PBA at CTE centers around the state. Verification of PBA testers and evaluators was completed during the month of March, and the first phase 2 PBA administrations were completed during the month of April.

The tests were administered to 550 students at CTE centers around the state during April of 2013. The course instructor and evaluators administered and scored the tests. Scores were collected via an online rubric module where test coordinators could input data. Data about the process were gathered via phone, email, and helpdesk logs, and a survey was given to instructors, evaluators, and CTE test coordinators after the testing was over.

Findings

Scores

The average PBA score was 81.78% with a pass rate of 90.36%. Comparing students' PBA scores to their Y1 averages reveals an increase of 17.02 points overall and a pass rate increase of 15.86%. See Table 1 for a breakdown of PBA scores by program.

Figure 1 shows a passing rate comparison between the Y1 and PBA scores. It is important to interpret this chart in light of Table 1, taking the number of students in each program into consideration. The number of testers has influence over the average PBA score because statistically averages become more accurate as the number of testers increases. For the purpose of this paper, we will consider only the data from Early Childhood Education as this was the largest group tested (n=354).

The Early Childhood Education (ECE) scores are a large enough data set that this program appears more statistically normal than the other programs. For example, in looking at a frequency histogram of the average difference between Y1 and PBA scores, ECE data returns an approximately normal distribution (bell curve). In Figure 2, an approximately normal grouping of score differences can be seen, from which we can conclude that there was a somewhat normalized pattern in the scoring of these tests.

The frequency of district average PBA scores can be seen in Figure 3, which returns a left-skewed distribution, showing that most ECE scores fell between 70% and 100%.

Usually, a bell-shaped distribution is desirable when looking at standardized test scores. However, if most students have mastered the course content, a more left-skewed distribution could be expected. The students who took PBA had been learning this content for at least one year, so the left-skewed distribution is less concerning than it would be had the data come from a random sample of untrained test-takers.

Problems with scores

Data were analyzed at the class level, meaning that they were organized into groups for which all subjects had been evaluated in the same setting and by the same instructor and evaluators. Three problems with PBA scores were found: abnormally agreeable scores, abnormally high scores, and abnormally irregular scores. Abnormally agreeable PBA scores occurred when the instructor and evaluators agreed completely on the scores for the performance scenario. This causes concern because while it is desirable for evaluators' ratings to agree with one another, scores with this level of agreement suggest that the evaluators discussed what the scores should be and influenced one another rather than rating students independently. In another setting, this type of

Brown, Miller

collaboration could be acceptable, but because one person (the instructor) is a stakeholder, this interaction allows for the instructor to influence evaluator judgment. There are four groups of scores for which this occurred and five groups for which half or more student scores showed perfect agreement. This problem affects approximately 19% of the scores.

Abnormally high and agreeable scores occurred when the teacher and evaluators agreed about each student's score and gave all students the highest score possible. This is not desirable as it is unlikely that all the students in one class performed at this level. Our data contains only one instance of this issue, representing 0.9% of the scores.

Abnormally irregular scores occurred when there was little agreement among the Part II scores. One explanation for results of this type is that evaluators did not completely understand what constituted each possible score. About 6% of all scores were abnormally irregular.

Reliability

A generalizability theory analysis was used to measure inter-rater reliability. This analysis returned a relative g -coefficient of .89 with the highest variance among students and the lowest variance among raters, a desirable pattern as one would expect students to perform according to their level of ability while raters should follow a more standardized pattern. Every criteria analyzed was found to be reliable at .87 or higher.

Survey Data

At the close of the testing window for PBA, a survey was sent to all test coordinators, instructors, and evaluators involved in the testing process. The aim of the survey was to gauge stakeholder perception of the PBA process and to uncover limitations of this spring's process in

order to improve for the next year. Sixty-four evaluators, 23 instructors, and 27 test coordinators responded with an overall response rate of 39.86%.

The survey revealed that many respondents felt PBA allowed students to be more creative. Several reported that PBA seemed to be a good way to assess knowledge since it was more similar to what the students do in class than a regular MS-CPAS2 test. The teachers enjoyed having a partial say in the students' scores. Test coordinators and evaluators appreciated that PBA allowed industry to see what CTE students are capable of doing. Overall, most respondents thought PBA was a good idea; however, they had concerns about the process which are detailed below.

Difficulty finding evaluators

It was initially thought that PBA evaluators would be easy to recruit from the community and the craft committee, a panel of individuals local to the CTE center who work in the industry into which the CTE pathway feeds. This was based on two assumptions: 1) programs are located in CTE centers based on industry needs in the area, and 2) CTE centers have established legitimate craft committees for each program. This proved to be true in several districts from which we received positive feedback about the process of recruiting PBA evaluators. However, qualitative data indicate this is not true for every CTE center. Test coordinators in several districts expressed frustration and concern about finding evaluators. In almost every incident, the RCU provided suggestions for where to find evaluators, and the problem was resolved quickly.

Despite the difficulties finding evaluators, 96% of test coordinators and instructors surveyed felt that evaluators were mostly interested and willing or very interested and willing to participate in the PBA process. Ninety-five percent of evaluators surveyed were mostly interested and willing or very interested and willing to participate in future administrations of PBA. Much of

the initial anxiety surrounding the recruitment of evaluators was related to the newness of the process, and it should become easier to recruit evaluators in future years.

Technology

PBA was delivered via Blackboard using Microsoft Word documents for the Early Childhood Education, Simulation and Animation Design, and Teacher Academy programs. A PBA demo was created for districts to try the technology ahead of time, and a backup plan of using printed materials was set in place for the districts that do not have the appropriate technology. The districts that utilized the PBA demo as a means of troubleshooting their technology were more prepared for PBA than the districts where the PBA demo was not used, and unfortunately, technology difficulties came as somewhat of a surprise to several districts. Issues included difficulty logging into Blackboard and trouble opening and saving files. The login issue was due to a scheduling setup where students could only access their PBA assignments during the time for which they were scheduled to test. If the district had not confirmed their Part II time with the RCU, materials were not available, and the RCU had to be contacted to make the materials available. Trouble opening and saving files may have been due to outdated software, and this is a problem that will have to be dealt with at the district level. The RCU will provide a technology recommendations list for future PBA as well as a larger emphasis on using the PBA demo to troubleshoot available technology and what to do if it is not working.

Survey data also indicated concern about students having limited knowledge of the technology used for PBA. All CTE centers in Mississippi have free access to a learning management system (LMS). In past years, the state provided access to Blackboard, and in the summer of 2013, Canvas by Instructure became the LMS of choice for Mississippi CTE. PBA for

2014 will be delivered via Canvas wherever possible, and recommendations will be made that instructors ensure that their students have time to learn how to use the system before taking PBA.

Evaluator leniency

Qualitative survey data indicates that some evaluators felt that the PBA tasks were too difficult for the students. A few evaluators made the comment that the performance scenario seemed to be written for college students. This demonstrates a lack of evaluator knowledge of the course curricula for the different PBA programs. This perception of difficulty may account for part of the high PBA scores as evaluators may have been more lenient when they felt a task was too challenging for a high school student. This can be corrected in part by including information about the CTE courses in the evaluator training so that the evaluators will recognize the level to which students should be achieving.

Test security

Test security for PBA allowed instructors both to look over the test materials and administer the test. This method of delivery asked teachers to view the materials before the test administration. Although the test coordinator, a proctor, or an evaluator also had to be present, there were two confidentially reported incidents of teachers being less than honest in regard to test security for PBA. The incidents involved teachers telling their students what was on the test ahead of time and/or attempting to influence the evaluator scores. The incidents were reported by test coordinators who offered the information on condition that their names were kept anonymous. Recommendations have been made for more stringent test security measures to prevent this situation in the future.

Site-wide implementation

Test coordinators were concerned about the amount of time it took to coordinate PBA for this year, especially in the larger districts. Several expressed that site-wide implementation would be very difficult. It is probable that some of this concern will decrease as the process becomes more familiar and changes are made in order to be as respectful as possible of human-resource constraints within districts.

Recommendations

Scalability

While an effective, working model for the Mississippi Career and Technical Education Performance-Based Assessment has been developed, scaling this type of assessment can be difficult. A plan for gradually implementing PBA throughout the CTE Pathways has been created, with approximately four programs scheduled to be added each year. Providing ongoing training for test coordinators, instructors, and evaluators will serve to smooth the transition to this type of assessment.

Training

Evaluators will be required to complete an online training course prior to serving as a PBA evaluator. The course will include information about the CTE program for which they are evaluating, a description of possible activities the students may be asked to complete, instructions for how to use the rubric, sample student responses to use as a guideline, and an overview of test security. Evaluators must pass the training in order to be qualified. PBA training materials will also be available for course instructors, test coordinators, test administrators, and proctors.

Technology

The 2014 PBA Test Manual published by the RCU will include a list of technology recommendations as well as a more detailed troubleshooting section. A larger emphasis will be placed on the importance of using the demo test to check for technological compatibility.

Test security

For 2014 PBA administrations, the course instructor will not administer Part II. The instructor will still interview students and score student work with evaluators; however, this will be done on a separate day. Collaboration between evaluators will not be allowed in order to ensure independence of students' performance ratings.

Time

The 2013 administration of PBA Part II consisted of the performance scenario and a brief session of questioning between the student and each evaluator. For 2014, Part II will be split into two parts. The new Part II will be a time for the students to complete a performance scenario, and Part III will be a 10-minute interview about the products created during Part II. Part III will be completed on a separate day from Part II in order to allow for plenty of evaluation time.

Conclusion

Performance-based assessment scores were higher than typical MS-CPAS2 assessment scores. Out of the four PBA programs, Polymer Science students scored the highest on average, and Polymer Science and Simulation & Animation Design students had the largest increase in scores between Year 1 and PBA. The score analysis indicates that as the number of students assessed by PBA increases, the overall scores tend toward a normal distribution. As more administrations occur, this principle will assist in determining validity and efficacy.

Many stakeholders initially displayed anxiety about the PBA preparation, testing, and reporting processes. This can be attributed to the lack of experiential knowledge of these phases of the PBA administration. At the conclusion of the spring 2013 administration, few participants surveyed still reported some form of anxiety or apprehension about continuing PBA in the future.

The instructors expressed an appreciation for the opportunity to play a role in the evaluation of their students. Results from the survey indicated that evaluators generally enjoyed gaining insight into the content taught in secondary CTE programs through this process. However, the score analysis from this spring's administration demonstrates the need for more rigorous evaluator training and tighter security measures. As this process continues in future administrations, more experienced evaluators and variation in evaluation procedures should produce more valid and reliable results regardless of the Career Pathway.

While there are details about PBA that require improvement, overall feedback about this assessment indicates that stakeholders believe PBA to be a good way to assess CTE students and that PBA can improve with time and training.

References

- American Psychological Association, Joint Committee on Testing Practices. (2004). Code of Fair Testing Practices in Education. Retrieved from <http://www.apa.org/science/programs/testing/fair-code.aspx>
- Darling-Hammond, L. and Adamson, F. (2013), Developing Assessments of Deeper Learning: The Costs and Benefits of Using Tests that Help Students Learn. Retrieved from the Stanford University, Stanford Center for Opportunity Policy in Education website: https://edpolicy.stanford.edu/sites/default/files/publications/developing-assessments-deeper-learning-costs-and-benefits-using-tests-help-students-learn_1.pdf
- Darling-Hammond, L. and Rustique-Forrester, E. (2005). The Consequences of Student Testing for Teaching and Teacher Quality. *Yearbook of the National Society for the Study of Education*, 104, 289-319. doi: 10.1111/j.1744-7984.2005.00034.x
- Lai, E. R., Wei, H., Hall, E. L., & Fulkerson, D. (2012). Establishing an evidence-based validity argument for performance assessment (White paper). Retrieved from Pearson website: http://www.pearsonassessments.com/hai/images/tmrs/Establishing_evidence-based_validity_argument_performance_assessment.pdf
- Lane, S. (2010). Performance assessment: The state of the art (SCOPE Student Performance Assessment Series). Retrieved from Stanford University, Stanford Center for Assessment, Learning, and Equity website: <https://scale.stanford.edu/system/files/performance-assessment-state-art.pdf>

Brown, Miller

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria (CSE Technical Report 331). Retrieved from University of California, Los Angeles, National Center for Research on Evaluation, Standards, & Student Testing website:
<https://www.cse.ucla.edu/products/reports/TECH331.pdf>

About the Authors

Ashley Priebe Brown is currently employed as a Project Manager and Data Coach by the Research & Curriculum Unit's Mississippi Assessment Center at Mississippi State University.

LeAnn Gupton Miller is currently employed as an Instructional Design Specialist by the Research & Curriculum Unit at Mississippi State University.

Table 1: Spring 2013 PBA Scores

PBA Program	Number of Programs	Number of Testers	Average Y1 Score	Average Y1 Pass Rate	Average PBA Score	Average PBA Pass Rate	Increase Between Y1 / Y2 PBA
ECE	35	354	65.11	73.43%	79.40	88.14%	14.29
PS	9	86	68.16	75.58%	90.85	97.67%	22.69
SAD	2	34	51.34	39.39%	76.18	88.24%	24.84
TA	10	76	74.42	93.42%	80.68	94.74%	6.26
Total	56	550	64.76	74.50%	81.78	90.36%	17.02

Figure 1: Passing Rate Comparison

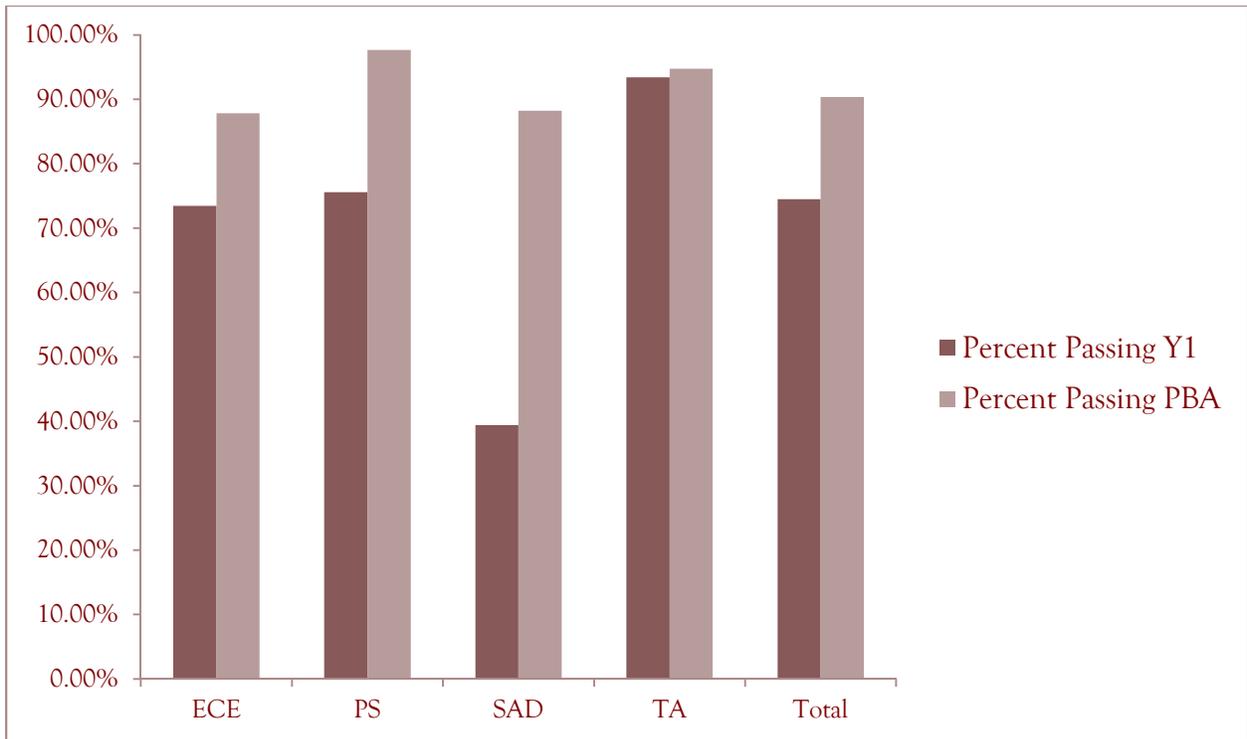


Figure 2: ECE Score Difference

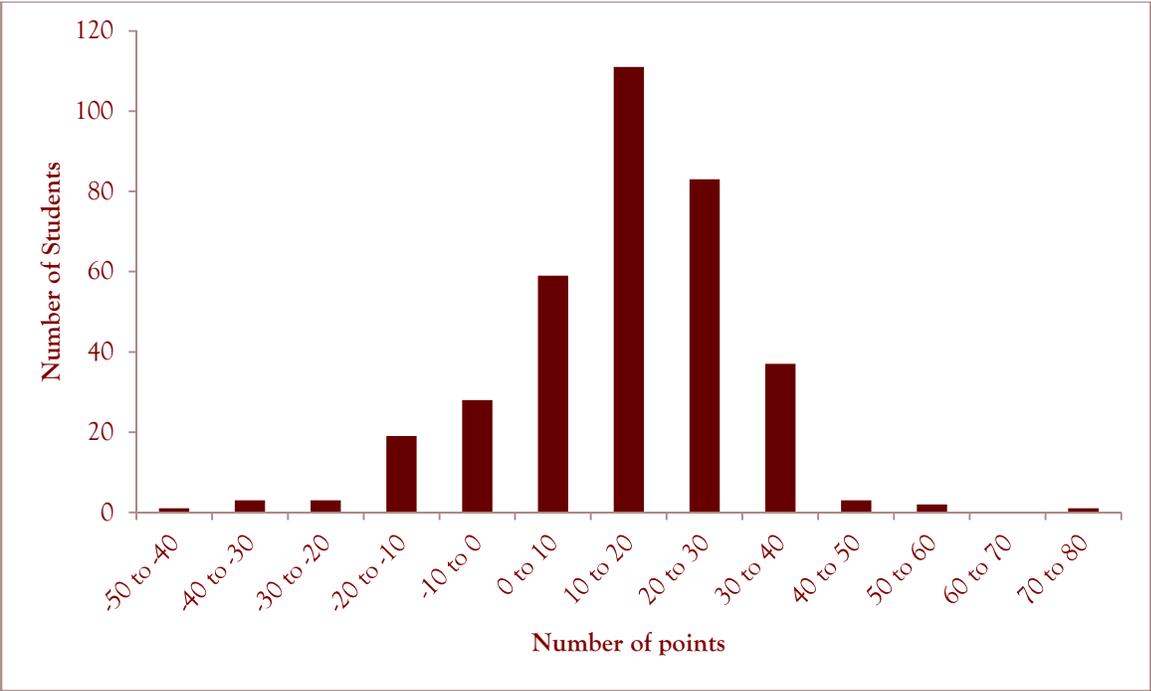


Figure 3: Frequency of ECE PBA Scores

