RSGC
Royal St. George's College

The Young Researcher

2017 Volume 1 | Issue 1

Comparing Traditional AP Classes to Portfolio AP Classes in Effectiveness of Measuring Knowledge

Leanne Olona

Recommended Citation
Olona, L. (2017). Comparing traditional AP classes to portfolio AP classes in effectiveness of measuring knowledge. *The Young Researcher*, 1(1), 176-193. Retrieved from http://www.theyoungresearcher.com/papers/olona.pdf

ISSN: 2560-9815 (Print) 2560-9823 (Online) Journal homepage: http://www.theyoungresearcher.com

Comparing Traditional AP Classes to Portfolio AP Classes in Effectiveness of Measuring Knowledge

Leanne Olona

High school students across the United States and internationally take Advanced Placement exams in May each year. These students also spend their school year enrolled in an Advanced Placement (AP) class in preparation for the exam. There is a lack of research on any correlation between a student's performance in class and performance on the exam. This study aims to compare the difference in correlation between traditional AP classes and portfolio AP classes. Second-semester grades and exam scores were collected from 2015 and 2016 from Norman High School students. Traditional classes had weak, if any, correlation, and portfolio classes presented no correlation. This lack of a relationship between class grade and exam score implies that students are unable to gauge their future exam performance based on class performance. Future researchers should compare data for a greater sample of students as well as regional samples.

Keywords: Advanced Placement, standardized testing, portfolio assessment, AP exams

Introduction

Almost all American high school students, parents, and teachers have heard about AP exams. Similar to honors classes, AP classes are designed to academically challenge students and offer potential college credit for those who perform well on AP exams. These classes are offered in all subject areas, English, Math, History, Science, as well as in Art and Music to name a few (AP Courses, n.d.). Most AP students hope to perform well on their exams and are likely to be disappointed when the result is not what they wanted. It is especially upsetting when a student receives a high grade in the class, but a low score on the test. This is because the relationship between class grade and exam score is unclear to students and parents. While there exists various research about AP exams, there is little to no current research over any connection between classroom grades and exam scores. On top of this, there is also a lack of research existing over the difference between traditional and portfolio AP classes. Traditional classes offer a single comprehensive sit-down exam at the end of the year in order to determine the AP exam score while portfolio classes decide the AP exam score based on a portfolio turned in by the student which consists of their work over the course of the year. To help take the first step to understanding the relationship between classroom performance and exam score, semester grade and exam score data from 2015 and 2016 school years at a high school in Oklahoma were compared to each other and the findings were discussed. While this sample size is only a small portion of the AP test-taking population, this research still provides new data for all members of the AP community. Standardized testing has been used for many different reasons to assess students, and

this research provides more insight into that conversation. In order to understand the results of the study, a discussion of current research precedes the results. Data were received from the local school district and are presented in graphs and tables. The results were discussed and a conclusion was reached based on the information provided.

Literature Review

As previously mentioned, almost all high school students in the United States are familiar with The College Board, an international testing company, and the Advanced Placement (AP) Testing that they offer. These high-stakes tests are designed to measure how much a student has learned in his or her AP class and scores them on a ranking from 1 to 5 (About AP Scores, n.d.). Colleges then determine how much credit they will give the student based on their own standards. Given that a 3 is a passing score, depending on the university, a 3, 4, or 5 can earn the student a certain amount of credit (About AP Scores, n.d.).

AP was created in 1953 by some of the top universities at the time, such as Harvard, Yale, and Princeton, with the initial intention to help high school seniors receive early college credit once they had completed their basic high school requirements (Fund, 1953; Casement, 2003). Today, AP classes are offered to all high school students, even freshmen. Students both domestically and internationally take AP tests every year, in addition to other standardized tests, such as the SAT and ACT. The growth of AP tests and other standardized tests have led to a discussion amongst researchers regarding the accuracy of these tests when it comes to measuring the knowledge of students.

Different Forms of Measuring Knowledge

In order to understand whether traditional AP classes or portfolio AP classes are more accurate in measuring what students know, it is important to look at various forms of testing measurements. While numerous types of testing exist, in regard to this research project, the most relevant forms to look into are IQ testing, GPA, Standardized Tests, and Portfolio Assessments.

IQ tests measure a person's "intelligence quotient". "IQ" is commonly used to refer to one's general intelligence, but there are, in fact, several different theories on types of intelligence; for example, Howard Gardner is well known for his ideas on multiple intelligences. The literature suggests that the most relevant form of intelligence in relation to standardized and portfolio assessments is successful intelligence (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Sternberg, 1988; Sternberg, & Kaufman, 2011; Sternberg, 2015). Robert J. Sternberg (2011) is the leading voice in the discussion on successful intelligence and he states that intelligence is an abstract concept and "not a quality of mind, but an aspect of behavior" (p. 26). Essentially, intelligence is less of a personality trait and instead reflects the behavior of an individual in a specific environment, such as a classroom.

Grade point average (GPA) is a number that comes from the average grades a student has received up to a certain point in their education. It is "assumed that grades reflect learning" so GPAs are reliable in that sense (Anaya, 1999, p. 500). However, others argue that problems with GPA include concerns of validity and generalizability (Anaya, 1999). GPAs do not reflect the growth of students and they cannot be compared across schools because grading may vary (Anaya, 1999).

Standardized Tests

Background. Standardized tests have been around since the 1800's. After the 1970s, they grew in popularity and became a normal aspect of the modern American student's education (Bures, Barclay, Abrami, & Meyer, 2013; Longo, 2010; Strauss, 2013; Sternberg, 2015). One example of the growth of standardized testing is AP tests. Based on data archived from AP for the United States, 566,720 students took at least one AP exam in 1997, while 2,538,998 students took an AP exam in 2016 (AP Data, n.d.). In less than 20 years, the number of test takers increased by over 300%, not counting international test takers.

Some might be surprised to learn that the format of standardized tests has stayed roughly the same over the course of two hundred years (Sternberg, 1988; Sternberg, 2015). The literature surrounding this topic suggests that psychological theory has been modified and enhanced through different generations, but this did not translate to testing (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Sternberg, 1988; Sternberg,

2015). Many psychologists have found that students perform better on assessments when they can be creative and explain what they know as compared to filling in bubbles on an answer sheet (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Sternberg, 1988; Sternberg, 2015).

While some aspects of standardized tests may be revised and changed over time, they remain consistent in how they are administered. The premise behind standardized tests is that they are reliable and as the name suggests, are standardized. Due to this, the testing companies that exist, such as Pearson and College Board, are very large bodies that make profits off the Scantrons students use to test (Gonzalez, 2012; O'Brien, Winn, & Currier, 2014; Pennolino, & Oliver, 2015). Pearson controls most state standardized testing across the United States very strictly; they require teachers to become certified by a Pearson test, provide the tests for students to take, and sell schools Pearson review guides to help teachers and students prepare for these Pearson tests (O'Brien, Winn, & Currier, 2014; Pennolino, & Oliver, 2015). Recently, College Board reformed their SAT so that the questions would be more accessible to testers (Gumbrecht, 2014). However, students are still taking an exam for at least three hours in order to receive a score based off their skill on multiple choice tests (Gumbrecht, 2014). This shows that overall, there has been no substantial change in standardized testing.

Disadvantages and advantages to using standardized tests. Standardized tests are often criticized for discriminating against minority students and students of low economic background, as well as for only measuring what a student has memorized and not what they actually learned (Anaya, 1999; Bures, Barclay, Abrami, & Meyer, 2013; Levine, 1985; Longo, 2010; Pennolino, & Oliver, 2015; Valencia, & Calfee, 1991). Many outside factors can also influence how a student performs on a test. For example, if a student suffers from test anxiety, they may perform poorly on their AP test, despite maintaining an A in the AP class (Levine, 1985; Longo, 2010; Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Valencia, & Calfee, 1991).

This is not to say that standardized testing only brings disadvantages to students, as there may be potential benefits associated with them. As previously stated, standardized tests have been around for many years. If used efficiently, they can help both state and federal governments judge how well a school is teaching its students (Goertz, & Duffy, 2001; Riffert, 2005). It is also important to remember that when AP tests were first implemented, they were only offered to high school seniors to help them get ahead in their college education (Fund, 1953; Casement, 2003). This initial plan was noble, and AP classes were the best way to accomplish this goal at the time (Fund, 1953; Casement, 2003). However, given that many years have passed and standardized tests have grown in popularity and prevalence, it is important to be open to new ideas in order to fix the problems currently in place in the system.

Portfolio Assessments

Portfolio assessments are not new, but they are used much less in comparison to standardized tests. The literature seems to suggest that portfolios are better at measuring what students actually learn in a class (Bures, Barclay, Abrami, & Meyer, 2013; Mills, 2009; Perry, 1998; Valencia, & Calfee, 1991).

There are different variations of portfolio assessments. Research on electronic portfolios shows that they are a reliable method for evaluating literacy in elementary students (Bures, Barclay, Abrami, & Meyer, 2013). There is, however, a lack of research over how electronic portfolios might be implemented for high school students and how they might differ from subject to subject.

Other forms of portfolio assessments include formative and summative assessments. Formative assessments compile information and work from a student's daily work, whereas summative assessments are compiled of chapter quizzes and unit tests (Mills, 2009). Regardless of which type of portfolio is used, they all provide a convenient method for teacher and student alike to measure their progress over time (Perry, 1998; Valencia, & Calfee, 1991).

A difficulty that arises out of the portfolio assessments is the subjectivity of grading. Portfolios are much more time consuming for teachers to grade since they are not tests that may be run through a machine, nor do they have keys (Mills, 2009). The individuality involved in portfolios makes the grading process much more subjective. Supporters of portfolio assessments claim that when teachers are prepared with the right rubrics, it can make the grading process

flow much smoother (Bures, Barclay, Abrami, & Meyer, 2013). Given the right resources, portfolios can be graded efficiently.

AP Class Credibility

A few voices have questioned the credibility behind AP classes. Because many universities, in America and across the world, give potential credit based on AP scores, two important core issues are validity and reliability (Credit & Placement, n.d.). The exams given to students must be accurate in their grading system in order to be considered valid, whereas all AP classes should be teaching their students the same information in order to be considered reliable. Some of the literature brings up how some Ivy League schools, Dartmouth specifically, have stopped giving credit for AP classes because students could not demonstrate mastery of a subject that they supposedly had passed; which suggests that AP is no longer being viewed as valid or reliable (Strauss, 2013). AP seems to lose some of its integrity when people have looked into the details of how they handle grading.

The literature suggests that with more people taking AP tests each year, the pass rate should be changing. Whether it should be increasing or decreasing is debatable, but there is a general consensus that there should be change; yet, the average pass rate has not changed much at all (Casement, 2003; Strauss, 2013). AP scores seem to be created with a scaling method; AP graders determine how well a student performed individually but AP itself will only give about the top 60% of students a passing grade (Strauss, 2013). This method of grading can also be seen in some states with standardized testing, where perhaps only a certain number of students are supposed to pass so graders are encouraged to hand out less passing scores, regardless of the actual work of a student (Gonzalez, 2012).

AP World History is a traditional AP class in which students read chapters in books weekly and take notes to later study for period exams. In May, students take the AP World History Exam and answer multiple choice questions along with a few essays. Regardless of the quality of work and effort they displayed throughout the AP class, they will receive potential college credit based on the AP exam score alone.

Norman High School AP Classes

At Norman High School, there are two portfolio AP courses; Studio Art and Capstone. Studio Art students build a portfolio of their best works of art and submit that for AP graders in May. Capstone is a two-year course that is made up of Seminar and Research classes (Seminar students take a traditional exam in May which affects a portion of their AP score). Seminar students begin their AP test in January and create a group research project, an individual project, and take a sit-down exam in May. Research students create one intensive research project and presentation to receive their AP score.

Despite all the differing opinions that can be read in the literature, there is a lack of research on AP in general, and no notable research has been done over high school students in Norman, Oklahoma. Since there are only two AP portfolio courses, there is also a lack of information about how valid and reliable their AP tests are.

Methods

The lack of statistical comparison between AP class grades and AP test scores presents an opportunity to discover new information that can potentially improve how students are tested. With the number of students taking AP tests increasing every year, it is important to check that these tests are both valid and reliable. A growing interest in portfolio programs in American schools may present an alternative to the traditional AP tests. Grades and AP exam scores were collected from Norman High School, located in Norman, OK. With district approval, the data was received and statistically analyzed in order to look for correlation.

Norman High School (NHS) is a strong candidate to conduct this statistical research for a few reasons. It has a large student body population, around 2,000 students, and offers 19 AP classes. In 2015, 819 AP tests were taken by NHS students enrolled in an AP class and in 2016, the number grew to 1005. The optimal way to approach the research was to use a quantitative method to look for a correlational relationship between end of class grades and AP grades. Students often have strong feelings about their grades so in an

attempt to avoid possible bias from surveying students, the best way to reach a conclusion was strictly through data.

The years 2015 and 2016 were selected because they are the most recent and because of the scope of the project; 2 years of data is the most that could be analyzed in the amount of the time provided. Data was also limited to NHS for this reason. While 19 AP classes are offered, data from AP Physics C E&M and AP Latin were not used since the students in those classes are engaged in self-study and have no more 10 students in each class. Students are allowed to take AP exams without being enrolled in the class, such as the AP Music Theory exam, which is irrelevant to this research and thus was excluded. Students in AEGIS Math and English take AP exams but since they are not enrolled in a traditional AP class and have a different class structure, their data is irrelevant to this research.

After receiving approval from the Norman Public Schools (NPS) institutional review board, an administrator within NPS provided me with the data for 1824 individual tests and corresponding semester grades.

AP teachers also were asked to provide a copy of their syllabus. This proved to be helpful in analyzing how a student received their end of semester grade in their course. Teachers were sent the request through an email and replied with attachments to their policies.

The data was first observed as a whole unit and separated simply by AP score. By approaching the data in this fashion, a general impression of the data was gathered before beginning a more in-depth analysis. This information was recorded and the data was then reorganized by traditional or portfolio. The end of semester grades was recorded, with A's changing to 95's, B's changing to 85's, etc. The results of the data analysis continue below.

Results

To consolidate the data collected, graphs for 2015 may be located in Appendix A. Similarly, graphs for 2016 may be located in Appendix B. The results section will refer to these graphs in terms of general trends, while providing figure number and appendix location for the reader. AP Exam Grade is represented on the x-axis, and Semester Grade is represented on the y-axis. The size of the bubble on the plot is related to the percentage of students who received the respective AP Exam Grade and Semester Grade.

Overall

With no regard to AP exam, class, or type of class, two graphs were created in order to observe general trends from 2015 (Figure 1, Appendix A) and 2016 (Figure 1, Appendix B). It can be seen that a majority of students received 3's on their AP Exams and A's in the class than any other combination of score and grade. It can also be noted that only a few students were able to score 5's on the exam without receiving an A in the class.

Traditional

15 exams fall under the traditional type as well as 14 AP classes. AP Comparative Government had just two students in 2015 and makes it incomparable to 2016 data. For this reason, the data is omitted.

AP exam performance aside, students generally received A's and B's in their classes. This may be seen in Biology, Calculus AB and BC, Chemistry, Environmental Science, Psychology, and Spanish Language for both 2015 (Figures 2, 3, 4, 5, 6, 9, 12, 13, Appendix A) and 2016 (Figures 2, 3, 4, 5, 6, 9, 12, 13, Appendix B).

Classes that showed greater variance in class performance fell under the subject of English, English Language and English Literature, or History, Human Geography, US History, and World History. Even in these cases, most students did earn A's in their classes in both 2015 (Figures 7, 8, 10, 15, 16, Appendix A) and 2016 (Figures 7, 8, 10, 15, 16, Appendix B).

Portfolio

There was a total of 5 portfolio exams taken by NHS students. Research and Seminar are from the two-year Capstone program, while Studio Art is broken up into 2D, 3D, and Drawing. Since AP Research has only has one year of data to look at, the data is omitted.

Except for Seminar grades in 2016 (Figure 17, Appendix B), all students in portfolio classes earned an

A or B in their class. Studio Art students in 2015 did not perform strongly on the AP exam, and scored 3's or lower in 3D (Figure 19, Appendix A) and Drawing (Figure 20, Appendix A). A few students were able to score a 4 in 2D (Figure 18, Appendix A). Students in Studio Art in 2016 followed this trend except that Drawing students scored 4's and a 5 (Figure 20, Appendix B).

Correlational Coefficients for 2015 for Traditional Exams	
Correlational Coefficient*	AP Exam
+0.49	Biology
+0.46	Physics C: Mechanics
+0.43	AB Calc.
+0.40	English Language
+0.40	U.S. History
+0.38	World History
+0.37	English Literature
+0.33	Chemistry
+0.32	Human Geography
+0.31	U.S. Gov. and Politics
+0.25	Psychology
+0.15	BC Calc. Subscore
+0.14	Environmental Science
+0.11	BC Calc.
-0.12	Spanish Language

* = Rounded to 2 decimal places

Correlational Coefficients for 2015 for Portfolio Exams	
Correlational Coefficient*	AP Exam
+0.53	Seminar
0.00	Studio Art: Drawing
-0.41	Studio Art: 2D
-1.00	Studio Art: 3D

* = Rounded to 2 decimal places

Correlational Coefficients for 2016 for Traditional Exams	
Correlational Coefficient*	AP Exam
+0.63	Environmental Science
+0.57	Human Geography
+0.51	U.S. History
+0.49	World History
+0.47	Biology
+0.47	English Language
+0.39	U.S. Gov. and Politics
+0.35	English Literature
+0.31	Psychology
+0.30	AB Calc.
+0.30	BC Calc.
+0.30	BC Calc. Subscore
+0.30	Physics C: Mechanics
+0.29	Chemistry
+0.22	Spanish Language

* = Rounded to 2 decimal places

Correlational Coefficients for 2016 for Portfolio Exams	
Correlational Coefficient*	AP Exam
+0.68	Seminar
+0.50	Studio Art: 3D
+0.20	Studio Art: 2D
0.00	Studio Art: Drawing

* = Rounded to 2 decimal places

Correlation

In order to explore the relationship between AP Exam score and class grade, a Pearson product-moment correlation coefficient was computed for each exam. The correlation coefficient, r, represents how strong the relationship is between two different numerical values. The closer r is to 1, the stronger the relationship is to being positive. The closer the r value is to -1, the stronger the relationship is to being negative. If the r value is 0 or close to 0, there is little to no relationship. Below are tables that contain the correlational coefficients for each exam.

Discussion

Before beginning a discussion of the results above, it is important to understand that correlation is not related to causation. The data collected above is a small sample of the entire population of AP test takers. Any conclusions that are reached in this discussion are limited and not definitive. This is only an observation of data since no experiment was conducted. The goal of this paper is to discuss the correlation between NHS student's AP scores and their final classroom grades.

Findings

It can be seen that in general, there is little to no correlation between AP Exam scores and Semester Grade. Traditional classes have the highest amount of correlation, while portfolio classes have almost no correlation. Most students in AP classes ended up receiving an A in their class which was expected since AP classes are intended for high achieving students. What stands out is that among students that earned A's in the class, most of them scored a 3 or lower on the exam, in both 2016 and 2015. The unique syllabi provided by AP teachers at NHS help explain why this may occur.

Traditional subject correlation. Traditional classes did have higher correlation than portfolios, but it was very weak. Since their grades include elements beyond testing, it is possible for students to obtain high final grades but not necessarily obtain high test grades during the semester.

Figure 3, in Appendices A and B, represents the data from AP Biology exams in 2016 and 2015. A majority of students finished their second semester with either an A or B in AP Biology but very few students were able to score a 5 on the exam. In the classroom, summative assignments account for 50% of the final grade, with five summative tests given in the second semester. Laboratory investigations and daily work each make up 25% of the final grade. For the student, this means that one poor exam grade will not have a major impact on their final grade, since they have four more exams and two other areas in the class in which they may succeed. Alternatively, the AP Biology Exam is split into two parts, multiple choice and free response, both of which make up 50% of the score. It appears that College Board has taken into account that students generally perform better when tests go beyond multiple choice, but half of the exam score still relies on a student's ability to bubble in answers (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Sternberg, 1988; Sternberg, 2015). All of these factors are possible reasons for why there is less than a +0.50 correlation between semester grade and exam score.

AP Physics C: Mechanics is represented by Figure 11 in both Appendices. The correlation coefficient was lower in 2016 than 2015, but both were less than +0.50, meaning that there was also a weak correlation between the semester grade and exam score. Students in the class must perform well in four different areas in order to have a high grade in the class. Daily work counts for 18%, homework is 20%, take-home tests are 25%, and in-class tests complete the 37%. While the average AP score and grade are usually a 3 and an

A, students in Physics C: Mechanics had more individuals score a 4 and A than any other combination. It is not clear as to why this varies from the overall average for both 2015 and 2016. Some possible reasons for this may be that the teacher taught the material well or the students were strong test takers, but further research must be conducted to be sure.

Portfolio subject correlation. Almost all of the students in portfolio classes ended the year with an A in the class. This is because the grades they received in the class were almost all participatory. The difficulty with grading portfolios presents itself here as teachers cannot help students gauge what their final grade may be because the teachers themselves do not know what the final product will be (Mills, 2009).

The AP Studio Art exam best exemplifies this trend. In all subsections of Studio Art, the class grade is based purely on if students turn in a piece of art each week, with no consideration for the quality of art. Figure 20 in both appendices represents the data from the Studio Art: Drawing exam and show that all students maintained an A in the class, yet had varying exam scores. In the year 2015, the highest exam score was just a 3. When testing for correlation, Studio Art: Drawing produced a score of 0.00. Most subjects had at least some level of correlation so it was surprising that this subject had none. This likely occurred because the weekly grades students receive in class are not representative of their final product. Participation grades will show no connection to assessment grades.

These findings support the problems universities like Dartmouth were having with students not showing complete mastery of subjects when they got to college (Strauss, 2013). Strong performance in the class does not guarantee a strong AP score and vice versa. In traditional and portfolio settings, it seems that students are not able to predict future AP test performance based on their class performance. This has heavy implications for all members of the AP community; as parents and students should be made aware that their in-class performance is a weak predictor of exam performance. Students that wish to excel on the exam will need to use other methods than work given by teachers. Teachers will find a very weak relationship between their assessment of students and College Board's assessment of the same students.

Alternative Explanations

As previously mentioned, the data collected came from a single high school. Since this is a sample of the population, it does not completely represent the whole group. It is possible that NHS employs several teachers who are not fit to teach AP, which would help account for the disparity between class performance and exam performance. Another possible explanation for the lack of correlation between the two variables is that students do not care about their AP exams. Class performance is associated with GPA while AP exams are a standalone test in which students face no real consequences for poor performance, so it is possible that students put more effort into the classroom and give up studying for the exams.

Limitations

This study is limited by several factors. It should be taken into consideration that data were requested from a single high school. Other factors may affect grades for students, such as test anxiety, socio-economic background, and any outside influences on a single student. Several AP classes at Norman High are taught by more than one teacher. Despite the teachers sharing a common syllabus, they each have their own unique teaching style. Not all students necessarily respond well to every teacher. To protect student privacy, all identifying information, including gender, ethnicity, and class rank were excluded from the data. These student demographics have proven before to be impactful on performance (Anaya, 1999; Bures, Barclay, Abrami, & Meyer, 2013; Levine, 1985; Longo, 2010; Pennolino, & Oliver, 2015; Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Valencia, & Calfee, 1991).

Class size also varied greatly between each subject. Data from several AP classes were not included due to class sizes averaging less than 4 people. Some of the small classes, such as AP Latin and AP Physics C: Electricity and Magnetism, are also independent study environments, which did not apply to this study.

There were also a few outliers in the data. A few students were able to earn 5's on their exams while having less than B's in the class. On the other end, students in AP AB and BC Calculus had the highest ratio of students scoring 1's on the exam, with A's in

the class. It should also be noted that AP Seminar was an outlier amongst the portfolio classes as it exemplified a positive, fairly consistent statistical significance. Due to the scope of this project, this data could not be analyzed in more depth.

Future Research

The number of limiting factors in this study provides opportunities for future researchers to continue exploring this topic. Future researchers should consider looking at a larger sample size. It is not feasible to attempt to gather the data for the entire population of AP students, but depending on the timeline of the researcher's project, they may analyze data from more than one school, as well as in other geographic regions. Students in different areas may possibly find more or less correlation between their semester grades and AP exam scores. Researchers are encouraged to seek out data from various years in order to look for a comparison over time. This would provide better insight to understanding the reliability of AP grades. As AP continues to improve and revise its curriculum for portfolio classes, future researchers will have many opportunities to compare these classes to traditional classes.

Conclusion

The continuous growth of students taking AP tests should be an encouraging force for future research into the accuracy of the tests and also into how students are being assessed in the classroom. While this study was limited to one high school, the weak relationship between class performance and exam performance, in traditional or portfolio subjects, raises concerns. Traditional standardized tests currently dominate the assessment industry but it is not certain they will forever. Portfolio AP tests are new and College Board, teachers, and students are all working to adjust to them. Further research into better implementation of portfolio based assessments is very important. AP tests are intended to prepare students in the best way possible for college, and should be held accountable.

References

About AP Scores. (n.d.). Retrieved June 03, 2017, from https://apscore.collegeboard.org/scores/about-ap-scores

Anaya, G. (1999). College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research In Higher Education*, 40(5), 499-526.

AP Courses. (n.d.). Retrieved June 14, 2017, from https://apstudent.collegeboard.org/apcourse

AP Data – Archived Data – Research – The College Board. (n.d.). Retrieved January 04, 2017, from https://research.collegeboard.org/programs/ap/data/archived

Bures, E. M., Barclay, A., Abrami, P. C., & Meyer, E. J. (2013). The reality of assessing 'authentic' electronic portfolios: Can electronic portfolios serve as a form of standardized assessment to measure literacy and self-regulated learning at the elementary level? / L'évaluation d'e-portfolio «authentiques». *Canadian Journal of Learning and Technology / La Revue Canadienne De L'apprentissage Et De La Technologie*, 39(4). doi:10.21432/t29g6w

Carpentier, M. (2016, August 09). No Child Left Behind has been unsuccessful, says bipartisan report. Retrieved September 20, 2016, from https://www.theguardian.com/education/2016/aug/09/no-child-left-behind-bill-unsuccessful-report-us-schools

Casement, W. (2003). Declining credibility for the AP program. *Academic Questions*, 16(4), 11-25.

Credit & Placement. (n.d.). Retrieved June 06, 2017, from https://apstudent.collegeboard.org/creditandplacement

Fund for the Advancement of Education. (1953). *Bridging the gap between school and college; A progress report on four related projects supported by the Fund for the Advancement of Education*. New York, NY: Fund for the Advancement of Education.

Goertz, M.E., & Duffy, M.C. (2001). Assessment and accountability systems in fifty states: 1999-2000. CPRE Research Report.

Gonzalez, S. (2012, June 4). Inside a 'scoring center' in the standardized testing industry. Retrieved November 03, 2016, from https://stateimpact.npr.org/florida/2012/06/04/inside-a-scoring-center-in-the-standardized-testing-industry/

Gumbrecht, J. (2014, March 06). Major changes coming to 2016 SAT test. Retrieved June 15, 2017, from http://www.cnn.com/2014/03/05/living/sat-test-changes-schools/

Levine, M. D. (1985). The measure of child. *Pediatrics*, 75(4), 791.

Longo, C. (2010). Fostering creativity or teaching to the test? Implications of state testing on the delivery of science instruction. *Clearing House*, 83(2), 54-57. doi:10.1080/00098650903505399

Mills, M. M. (2009). Capturing student progress via portfolios in the music classroom. *Music Educators Journal*, 96(2), 32-38.

O'Brien, J., Winn, R., & Currier, K. (2014). Not all tests are created equal: Parental rights and standardized tests. *Journal Of Cases In Educational Leadership*, 17(2), 41-48.

Pennolino, P. (Director), & Oliver, J. (Producer). (2015, May 3). *Standardized Testing* [Video file]. In YouTube / Standardized Testing. Retrieved October 23, 2016, from https://www.youtube.com/watch?v=J6lyURyVz7k&t=889s

Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal Of Educational Psychology*, 90(4), 715-29.

Riffert, F. (2005). The use and misuse of standardized testing: A whiteheadian point of view. *Interchange: A Quarterly Review Of Education*, 36(1-2), 231-252.

Rotherham, A. J. (2012, January 6). In defense of No Child Left Behind. Retrieved January 04, 2017, from http://ideas.time.com/2012/01/06/in-defense-of-no-child-left-behind/

Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a basis for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology*, 31(3), 344-376.

Sternberg, R. J. (1988). applying cognitive theory to the testing and teaching of intelligence. *Applied Cognitive Psychology*, 2(4), 231-255.

Sternberg, R.J., & Kaufman, S.B. (2011). Tests of intelligence. In *The Cambridge Handbook of Intelligence* (pp. 20-38). Cambridge: Cambridge University Press.

Sternberg, R. J. (2015, October 1). Successful intelligence: A model for testing intelligence beyond IQ tests. *European Journal of Education and Psychology*, 8(2), 76-84. doi:10.1016/j.ejeps.2015.09.004

Strauss, V. (2013, February 22). Teacher questions value of AP program. Retrieved November 03, 2016, from https://www.washingtonpost.com/news/answer-sheet/wp/2013/02/22/teacher-questions-value-of-ap-program/

Valencia, S. W., & Calfee, R. (1991). The development and use of literacy portfolios for students, classes, and teachers. *Applied Measurement In Education*, 4(4), 333-45.

Appendix A

There was a mix of numerical and categorical variables involved in the collection of data. AP Exam Grade refers to the score that a student received on their AP exam, ranging from 1 to 5. Semester Grade refers to the final grade that a student received in their AP class, ranging from 55 to 95. There was a total of 20 different AP exams, within 17 AP classes. AP Government is a yearlong class at NHS, but there are two separate exams, Comparative Government and U.S. Government. AP Studio Art is also offered to students at NHS for a year, but within the class students choose to take the 2D, 3D, or Drawing AP exam. The AP BC Calculus exam provides students with two scores; their BC score and AB subscore. This data was presented in two separate graphs. These 20 AP exams were also split up by type, either traditional or portfolio. Data is represented from 2015.

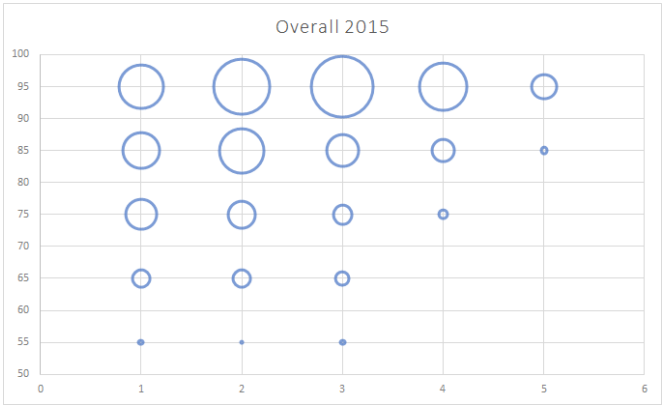

Figure 1. This graph presents an overall view of all the data collected from the year 2015. Most students obtained an A in the class, and a 3 on their exam.

Traditional Classes

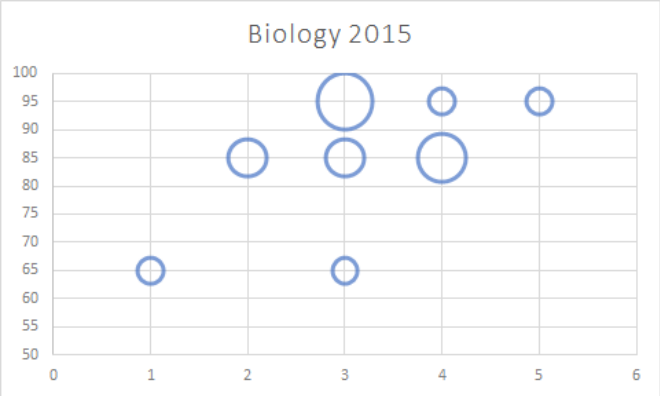

Figure 2. Biology students in 2015 mostly finished their second semester with an A or B in the class. Two outliers ended the year with D's. AP exam score varied greatly.

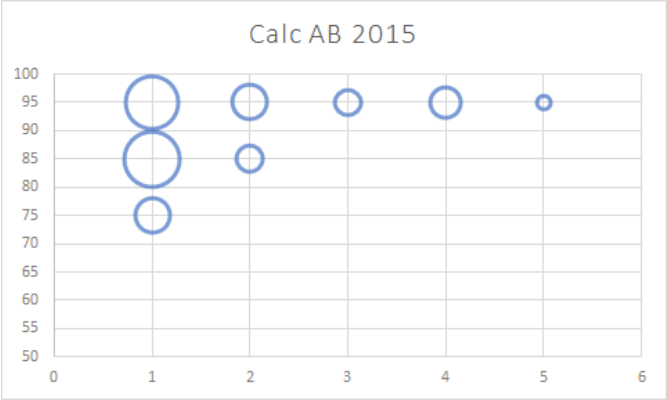

Figure 3. Students in Calculus AB had a majority of students perform poorly on the AP exam, despite earning A's and B's in the class. Only students with A's passed with a 3 or higher.

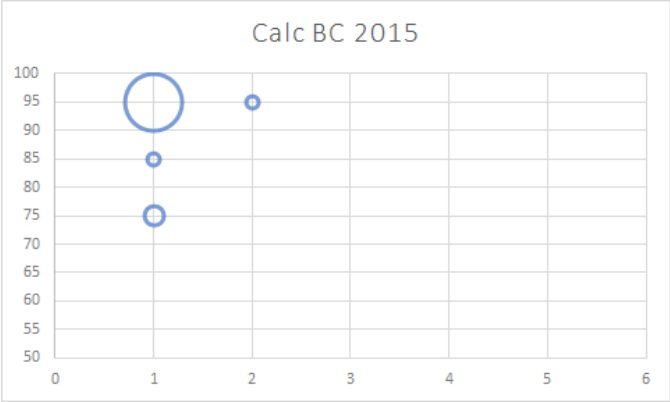

Figure 4. Students in Calculus BC performed very poorly on the AP exams in 2015. While a majority of students completed the year with an A in the class, no students passed the AP exam.

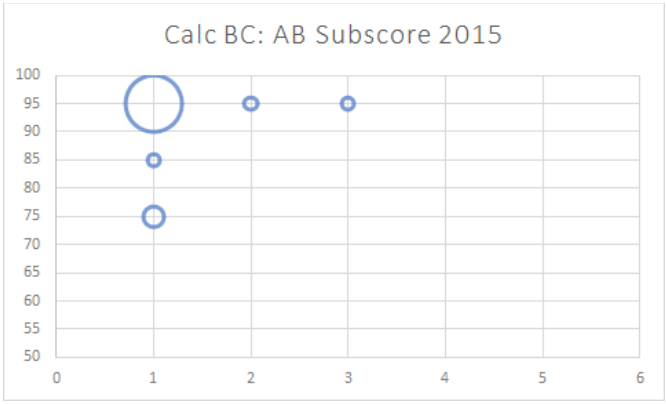

Figure 5. The Calculus BC exam is given two scores. This graph represents the AB subscore for students. Only one student passed this exam.

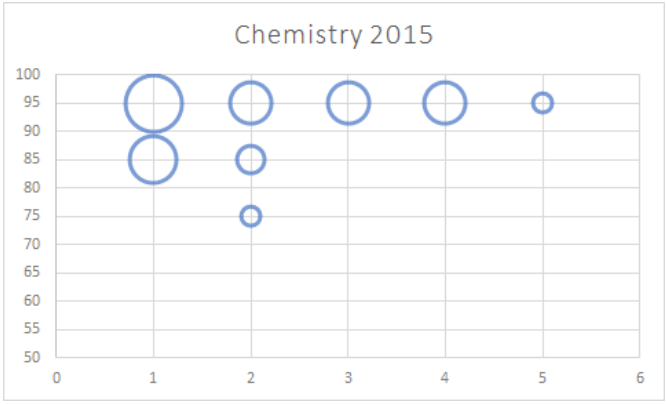

Figure 6. Chemistry students generally performed well in the class setting. Most students scored 1's on the exam, but many also scored 2's, 3's, and 4's. One student scored a 5.

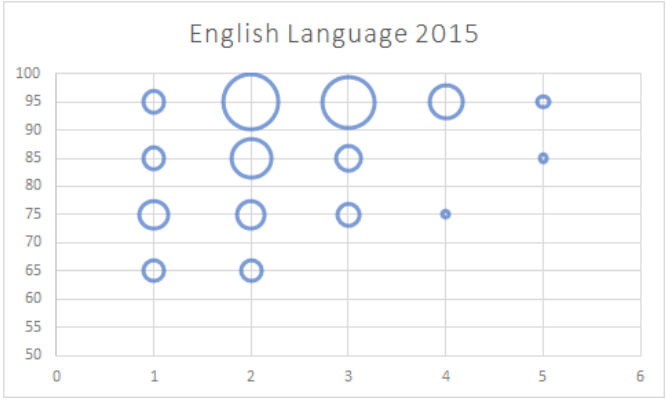

Figure 7. There was high variety between class grade and exam score in English Language in 2015. Most students earned A's in their class and scored 2's and 3's on the exam. A few students earned 5's on their exam as well.

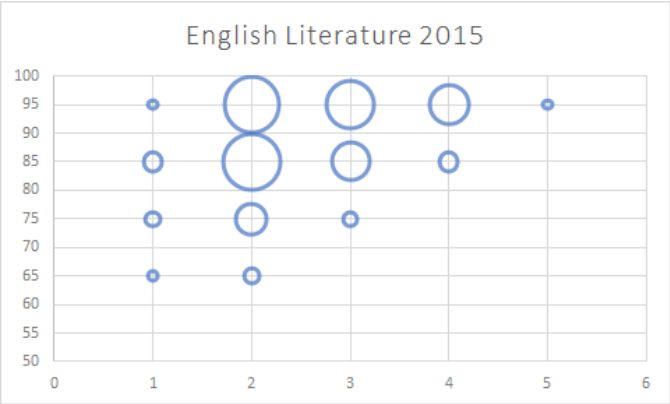

Figure 8. English Literature had very few students who scored 1's and 5's. Most students earned 2's and 3's.

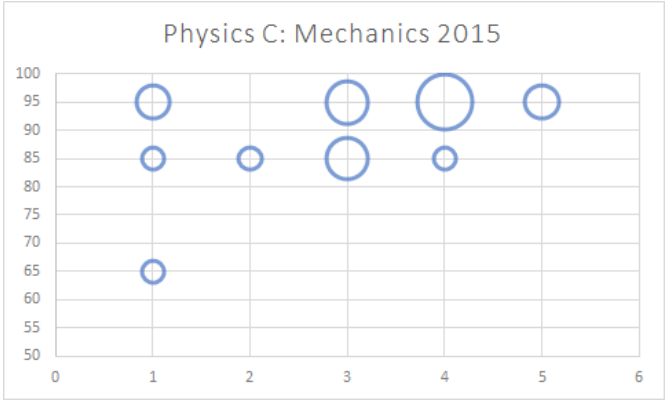

Figure 11. Almost all Physics C: Mechanics students finished the year with an A or B in the class. Most students earned 4's on their exams as well.

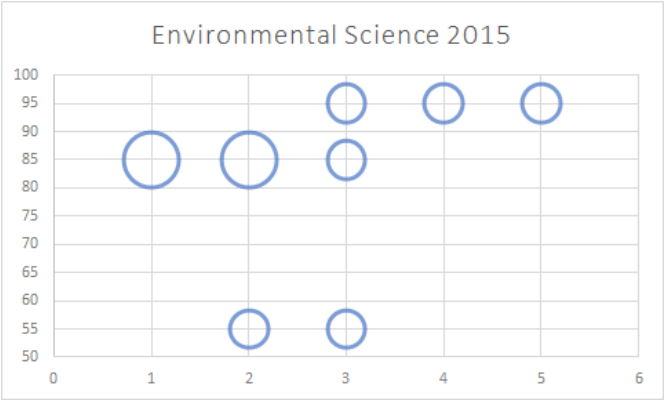

Figure 9. Students in Environmental Science either did well in the class, with A's and B's, or failed. It can be seen however that the only students with 1's on the exam finished the class with a B.

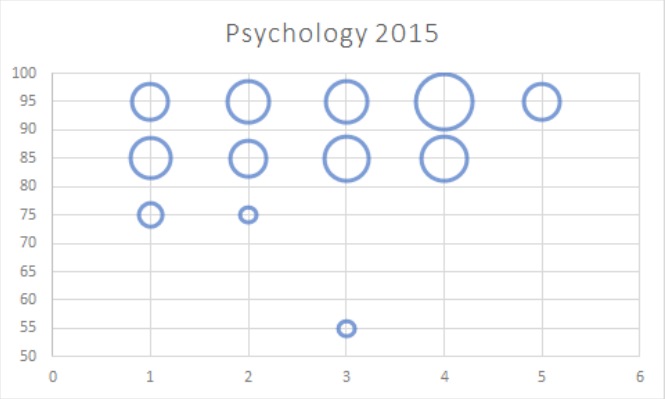

Figure 12. All students in Psychology, except for one outlier, ended 2015 with at least a C in the class. AP exam scores were much more dispersed.

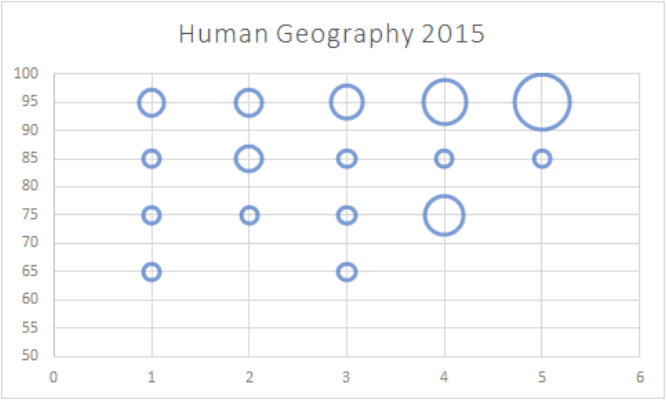

Figure 10. Exam score and class grade were extremely dispersed for Human Geography students. A majority did end up earning 5's on the exam and an A in the class.

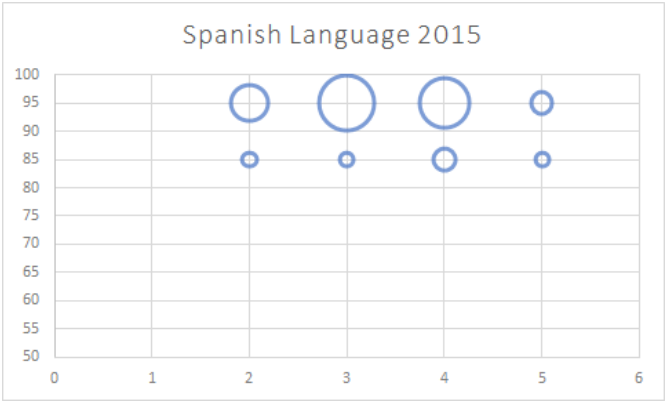

Figure 13. All students earned B's and A's in their Spanish Language class in 2015. Almost all passed the AP exam too, but a few students did receive 2's.

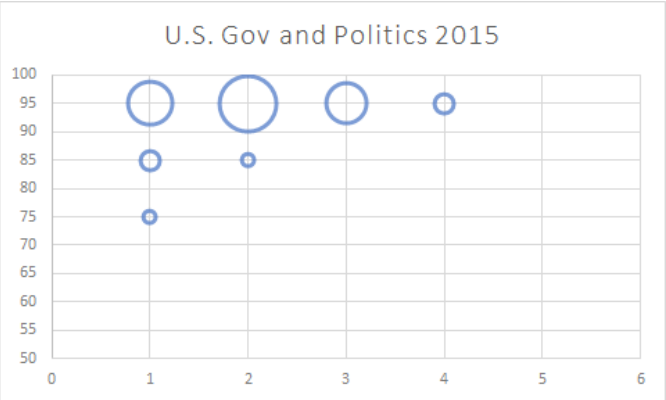

Figure 14. Students taking Government performed well in class, but not as well on the AP exam. Most students earned 2's, or a 1 or 3.

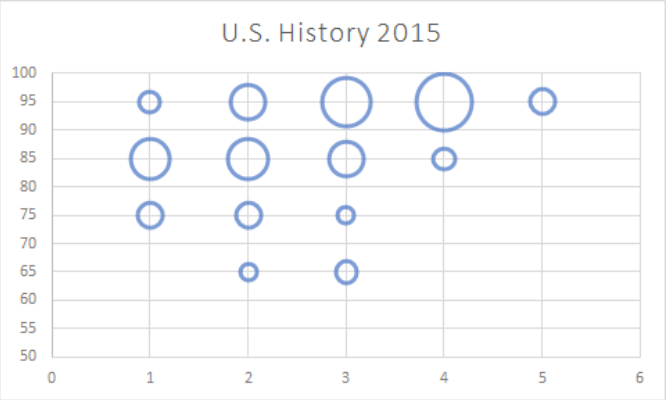

Figure 15. Grades and exam scores were varied greatly in U.S. History. A majority of students did earn both an A and 4. Several students were also able to obtain 5's on their exam as well.

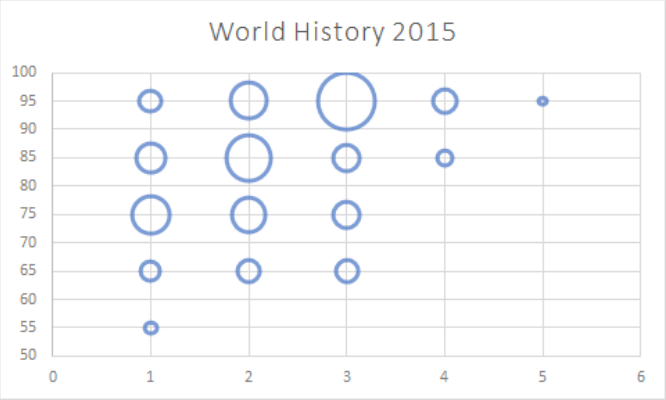

Figure 16. World History students generally did not pass their AP exam, or passed with a 3. Few students earned 4's, and only one earned a 5.

Portfolio Classes

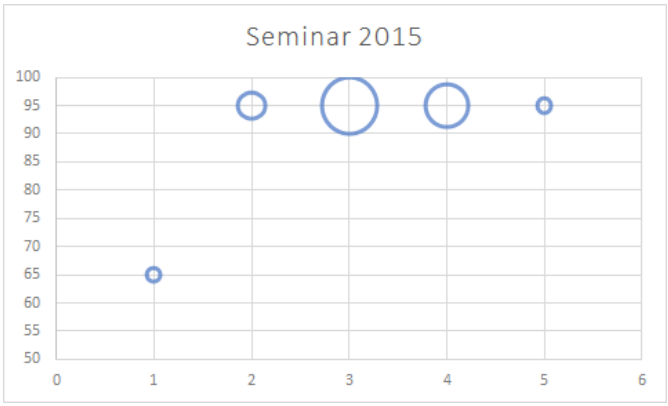

Figure 17. Almost all Seminar students finished the year with an A in the class. Most students earned 3's on their exam, while several achieved a 4.

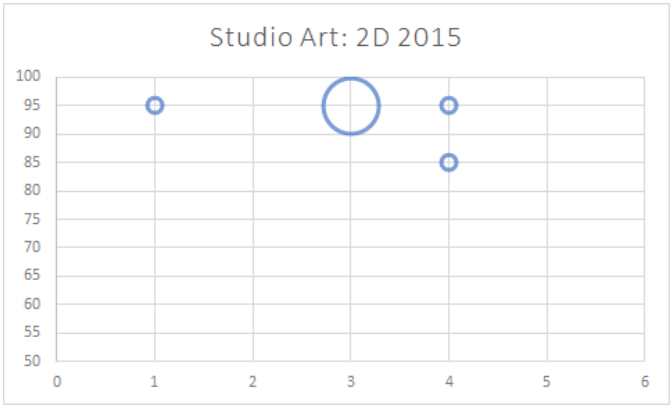

Figure 18. All students earned an A in the class and 3 on the Studio Art: 2D exam except for 3 outliers.

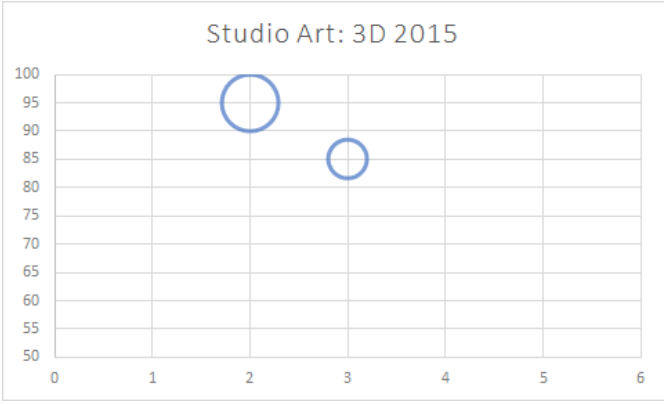

Figure 19. Students in Studio Art: 3D either had an A in the class and 2 on the exam, or a B in the class and 3 on the exam.

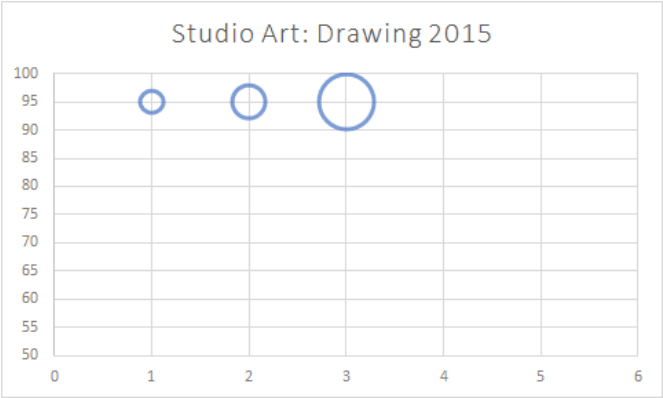

Figure 20. All Studio Art: Drawing students earned A's in their class. On the AP exam, students only scored 3's or lower.

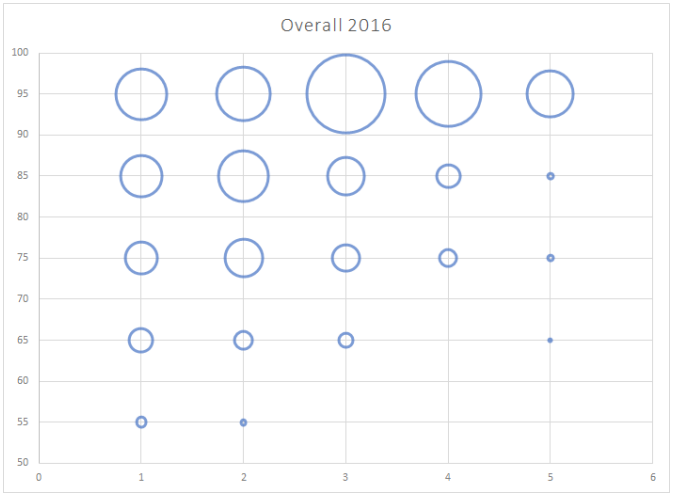

Figure 1. This graph presents an overall view of all the data collected from the year 2016. Most students obtained an A in the class, and a 3 on their exam.

Appendix B

There was a mix of numerical and categorical variables involved in the collection of data. AP Exam Grade refers to the score that a student received on their AP exam, ranging from 1 to 5. Semester Grade refers to the final grade that a student received in their AP class, ranging from 55 to 95. There were a total of 20 different AP exams, within 17 AP classes. AP Government is a yearlong class at NHS, but there are two separate exams, Comparative Government and U.S. Government. AP Studio Art is also offered to students at NHS for a year, but within the class students choose to take the 2D, 3D, or Drawing AP exam. The AP BC Calculus exam provides students with two scores; their BC score and AB subscore. This data was presented in two separate graphs. These 20 AP exams were also split up by type, either traditional or portfolio. Data is represented from 2016.

Traditional Classes

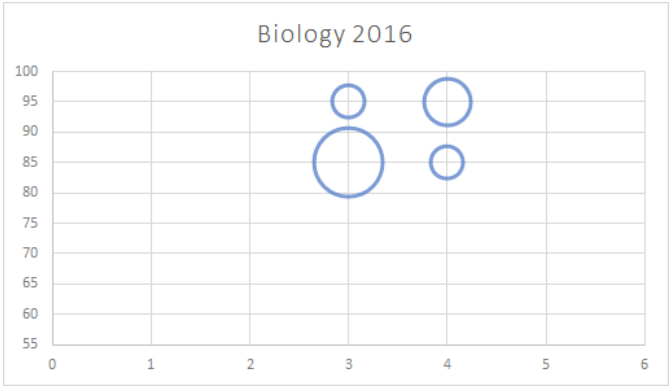

Figure 2. No Biology students earned less than an A or B in 2016. All students also scored a 3 or 4 on their exam.

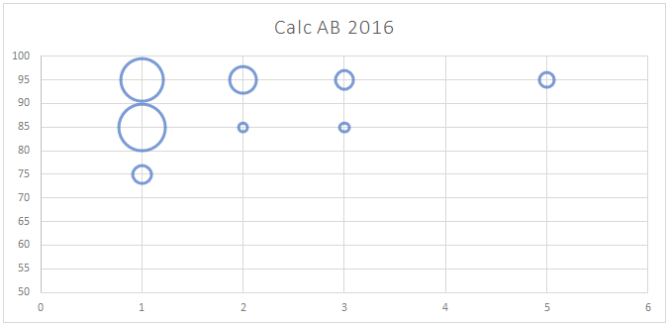

Figure 3. A few outliers were able to obtain 5's on the Calculus AB test, but most students scored a 3 or lower. All students passed the class though with a C or higher.

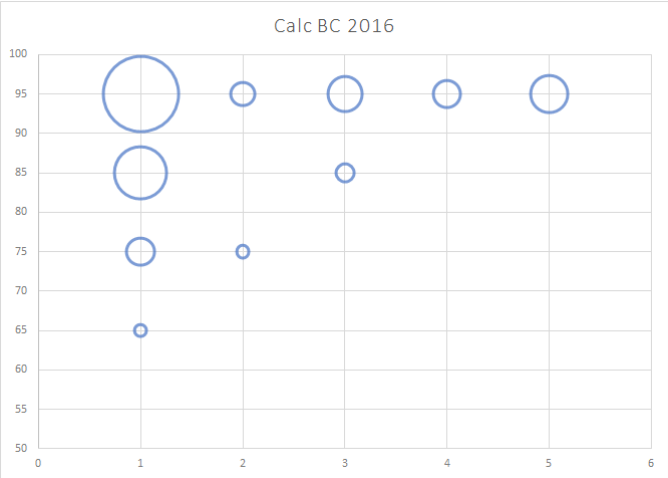

Figure 4. Calculus BC students had varying degrees of performance on the AP test. A majority did not pass, and ended up receiving 1's. Students who did pass almost equally earned 3's, 4's, and 5's.

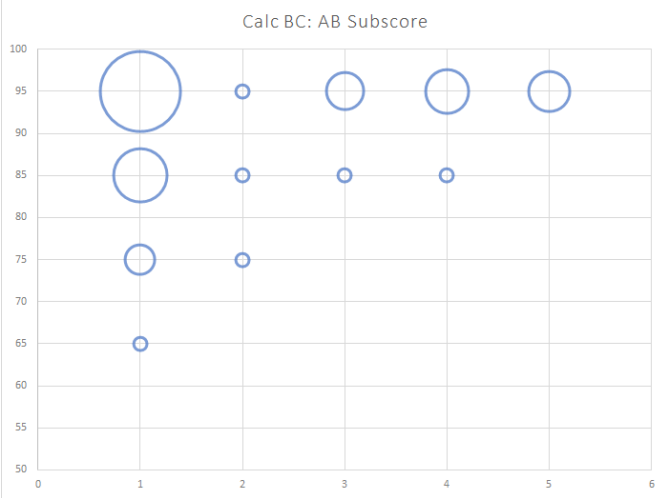

Figure 5. The Calculus BC exam is given two scores. This graph represents the AB subscore for students. Most students also received 1's on this portion of their AP test as well.

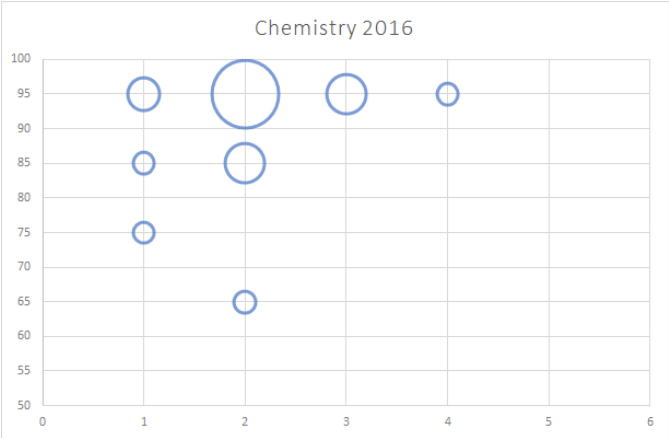

Figure 6. Most Chemistry students scored 2's on their AP exam, but many finished the class with an A.

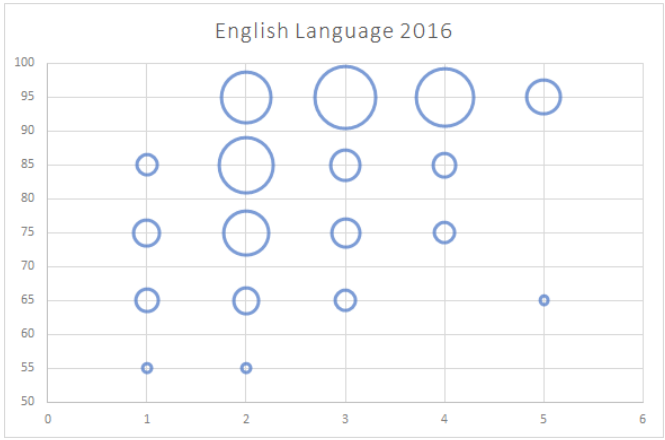

Figure 7. Students performed very differently on the English Language exam. Many students did earn A's in their class and a 3 on the exam. One outlier earned a 5 and had a D in the class.

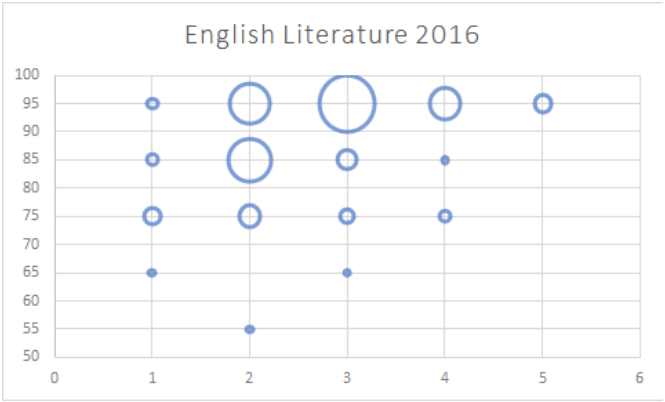

Figure 8. Several students earned 1's on the English Literature exam, despite their class grade. Most students had A's in the class and scored 3's on their exam. A few were able to earn 5's.

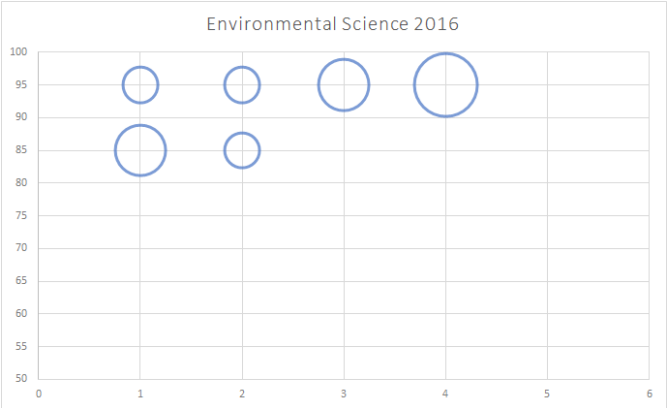

Figure 9. All students in Environmental Science finished the year with at least a B in the class. Half of the class failed the exam while the other half passed.

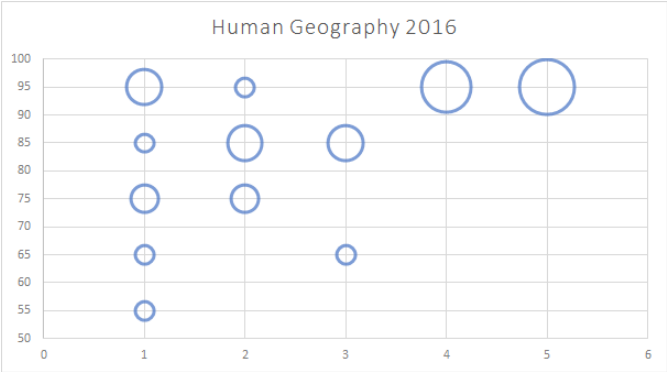

Figure 10. Students scored 1's on the Human Geography exam while also finishing the class with every type of final grade. Several students did earn 4's and 5's, with A's in the class.

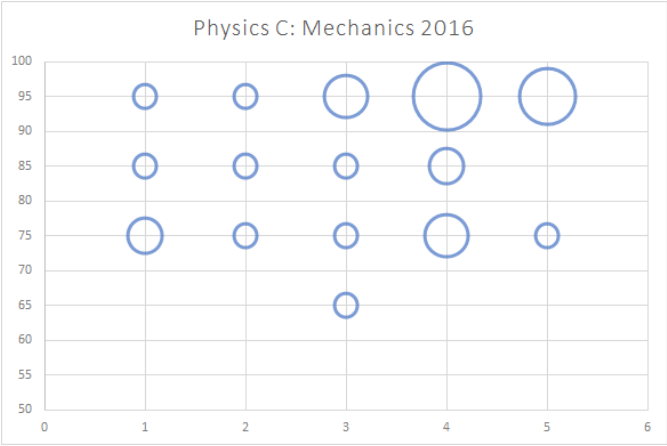

Figure 11. Almost all Physics C: Mechanics students finished the year with an A or B in the class. Most students earned 4's on their exams as well.

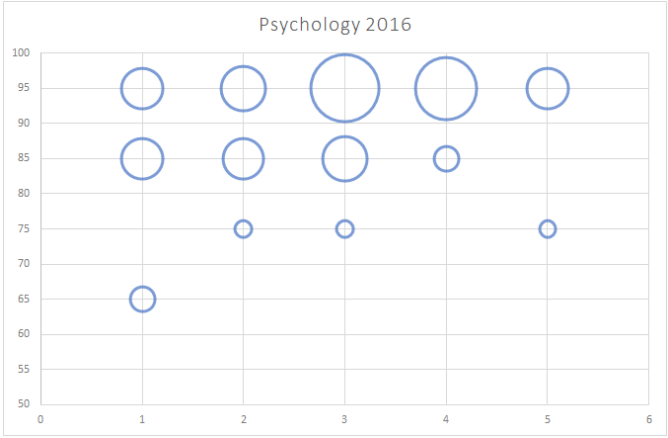

Figure 12. A majority of Psychology students earned A's in the classroom. Several students obtained 5's on the exam, with one outlier who had a C in the class.

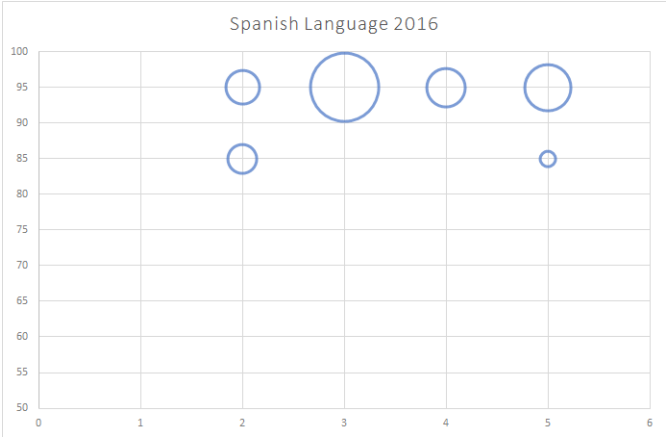

Figure 13. All students earned B's and A's in their Spanish Language class in 2016. Most students got a 3 on their exam but several received 5's. An outlier had a B and a 5.

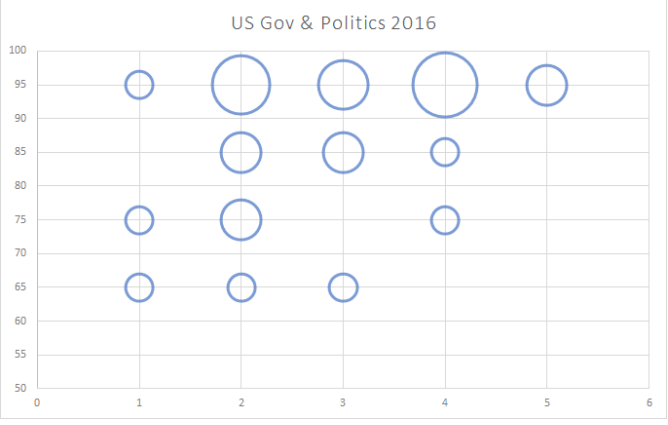

Figure 14. There was great variance between all combinations of class grade and exam score. Most students either had an A and a 4, or an A and a 2.

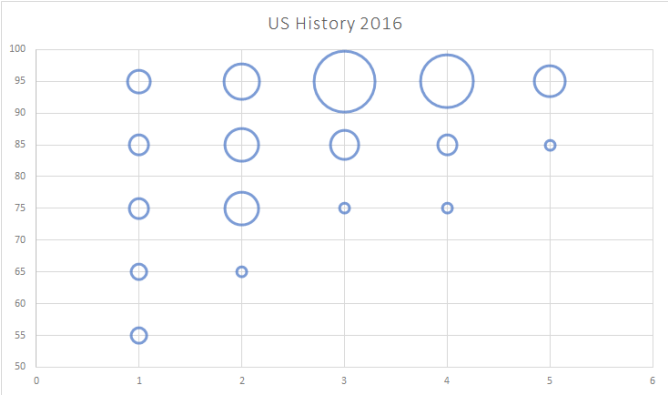

Figure 15. Grades and exam scores were varied greatly in U.S. History. A majority of students did earn both an A and 3 or a 4. Several students were also able to obtain 5's on their exam as well.

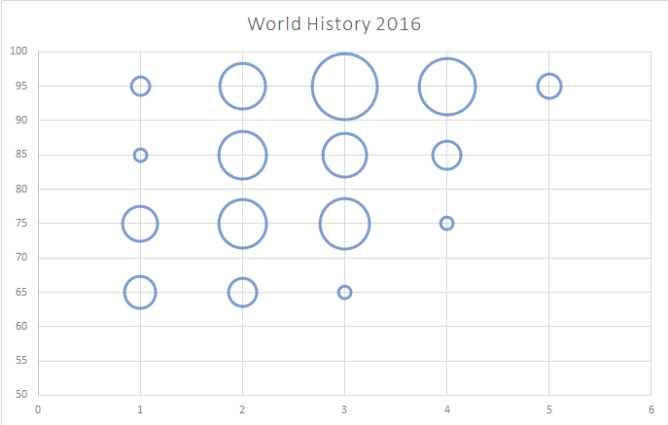

Figure 16. World History students generally did not pass their AP exam, or passed with a 3. Few students earned 4's, and several earned a 5.

Portfolio Classes

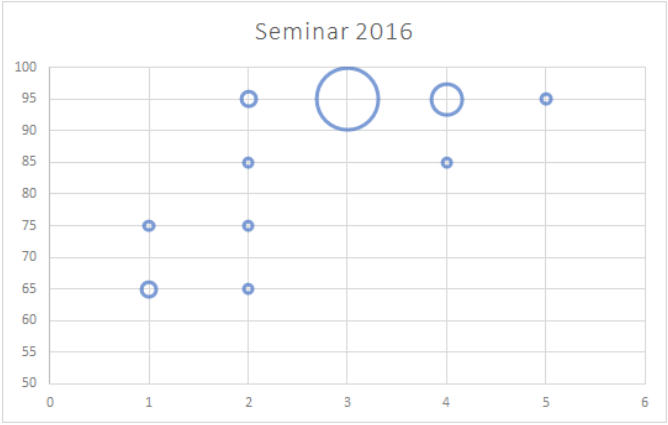

Figure 17. 2016 Seminar students had many outliers. A majority of students earned 3's and A's, and several 4's also occurred. The other combinations were almost all individual.

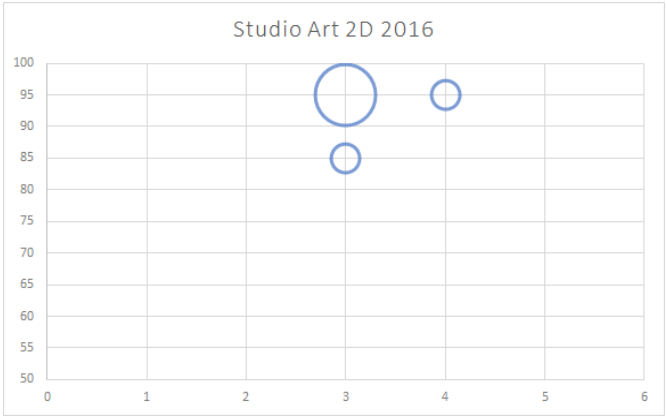

Figure 18. A majority of students earned an A in the class, as well as a 3 on the exam. Two students experienced other combinations of scores.

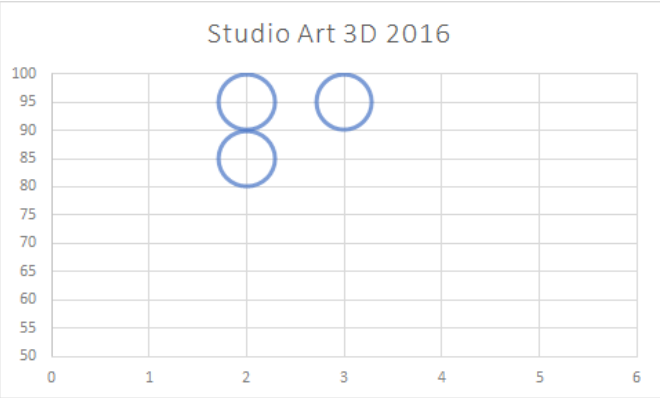

Figure 19. Students in Studio Art: 3D either had an A in the class and 2 or 3 on the exam, or a B in the class and 2 on the exam.

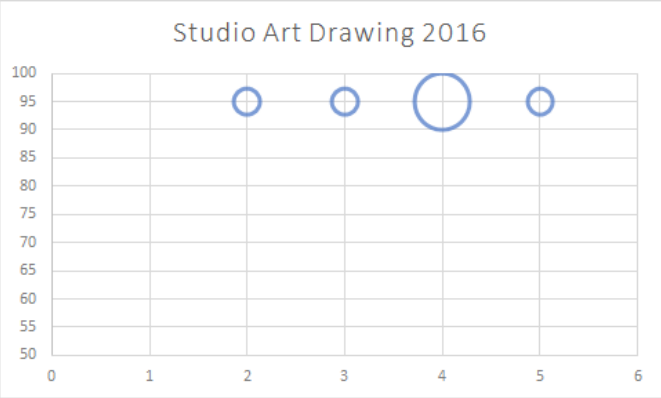

Figure 20. All Studio Art: Drawing students earned A's in their class. On the AP exam, students mostly scored 4's.