



## Peer Assessment in AP Government and History Classes By David Steere

### Introduction

Grades play an outsized role in American education, both inside and outside educational institutions. From nearly the beginning of American education, grades have been used to sort students. But despite the seeming clarity of a single end-of-semester score, educators have raised numerous questions about the grading process. Grading accuracy and giving timely and helpful feedback have been challenges for educators according to academic research and also confirmed by my own classroom experience.<sup>1</sup> Teachers are often overwhelmed with work and cannot return longer assignments in a timely manner. Students sometimes feel that the grading process is not fully transparent, leading to anxiety, especially given the pressure many students feel during the college admissions process. Others posit the potentially demotivating impacts of grading as well as the pervasiveness of cheating as a way to cope with the high stakes nature of grades.<sup>2</sup> Grading also gives teachers enormous power, which can create discomfort within the student-teacher relationship and feelings of helplessness among students. Indeed, even discussions about grading are often seen as taboo, like a “train’s third rail.”<sup>3</sup> As Feldman points out in *Grading For Equity*, scrutiny of grading practices for both students and teachers can invoke “anxiety, obstinacy, insecurity, pride and conflict.” In the face of these massive challenges, some have even proposed abolishing grades altogether.

Addressing all of these challenges can be difficult, if not impossible. To ameliorate some of the potential problems, educators should strive to be more transparent and honest about these challenges. This would help manage the emotional element of learning for both students and teachers. Open communication may also build trust between teachers and students, which could decrease cheating. More transparency about grading may also lead to greater student confidence. Peer assessment offers a possible way to create an atmosphere where grading is more transparent, less arbitrary, and more empowering for students. Peer assessment could also offer a way for students to receive feedback in a more timely manner since a single teacher would not be burdened with grading a large number of essays.

---

<sup>1</sup>Guskey 35.

<sup>2</sup> Alfie Kohn.

<sup>3</sup> Feldman.

## Peer Assessment

AP classroom teachers often have the opportunity to attend professional development workshops where teachers learn how to apply College Board rubrics to sample essays. This type of training is very popular with teachers because it helps them discern which responses earn points and which do not. Grading accuracy can improve with practice; indeed, it must because essays have to be scored in as uniform a manner as possible on AP exams. During these training sessions, teachers feel comfortable comparing their way of thinking about writing with other teachers. Looking at a wide range of student samples builds confidence for teachers and exposes teachers to many common mistakes students can make and how to recognize them. In addition, the AP rubrics discipline teachers into grading efficiently and not agonizing over any one essay.

While this process is generally restricted to teachers, students, too, could benefit from such training through peer assessment. The literature on peer grading is extensive. It has been used in all subjects and educational levels, from grade school to graduate level courses, and it has numerous manifestations. For example, in some cases, peer grading has been done throughout the school year while in other cases it's applied on only one assignment. Sometimes students interact with each other while peer grading while in other cases there is limited dialogue between students. Some educational fields have seen numerous peer grading studies while others have seen far fewer. But key questions have emerged over the efficacy of the peer assessment, which I attempted to address in my own exploration of this topic.

## Literature Review of Peer Assessment

Grading accuracy has been a chief concern of the peer grading literature. Can students grade as accurately as professional educators? Keith Topping has found that, generally speaking, peer grading is accurate, though the research on whether students can improve their grading skills over the course of year is limited.<sup>4</sup> Sometimes, student graders cluster scores around the mean when grading assessments, an observation not limited to amateur graders only, but his research has indicated that students grade about as well as professionals most of the time. In terms of assessment areas, patterns have not emerged across subject matter, with some subjects (life sciences) showing better accuracy than others (psychology and engineering classes).<sup>5</sup> Students may grade more harshly or leniently depending on the class. Stronger students also, on average, generally don't grade much more accurately than weaker students, except in the cases of extremely strong or extremely weak students.<sup>6</sup> Some research has found that student grading

---

<sup>4</sup> Topping 275.

<sup>5</sup> Freeman and Parks 486.

<sup>6</sup> Alfaro 69.

accuracy may decline with more complex grading tasks that require students to evaluate higher order thinking skills.<sup>7</sup> Students grading in teams also does not appear to improve performance.<sup>8</sup> Finally, higher performance students may suffer if their peer graders feel a need to mark them lower as part of the project.

In addition to grading accuracy, the literature also raises the question of how helpful the feedback is to the student receiving the peer assessment. For both peer graders and professional educators, giving the correct kind of feedback is important yet notoriously difficult, as students may not read the feedback carefully or fully understand it. Feedback can be too general or vague, even reinforcing underachievement if graders are not sufficiently honest. Some studies have found that students do not even read the feedback, let alone process it and then act on it during future assignments. Some teachers have seen that unless the feedback is helpful, even if the grading itself was accurate, peer grading can be problematic. For example, some students may simply write unhelpful comments such as “good job” or do not know how to give constructive feedback. They may simply lack the vocabulary to do so. Other students may feel overwhelmed and not know where to start with giving feedback, especially if the work is of very poor quality.<sup>9</sup> Students may also not want to hurt the feelings of their peers. Such feelings are perfectly understandable since students have been shown to be sensitive to the comments they receive from peer graders, even if the feedback is given anonymously. Students may also feel intimidated if they are asked to assess superior quality work that they feel is out of reach for themselves. Various studies have also raised the question of anonymity in their work. Will students fear some type of retaliation if the writer knows who graded their work? Research indicates, however, that anonymous feedback does not substantially impact the accuracy of peer grading.<sup>10</sup> In addition, non-anonymous grading limits the possibilities for collaboration and oral feedback.

Some of the literature suggests evaluating the graders to encourage accuracy and thorough feedback. One study apportioned the following points to a peer grading assignment: 70% for the student’s own assignment, 15% for the accuracy of the student’s grading of a peer’s assignment, and 15% for the helpfulness of the feedback. This approach may help students take the assignment more seriously.<sup>11</sup> Some educators also advocate breaking down peer grading activities into small chunks so students just evaluate, for example, one part of an essay; they may also advocate further evaluation even after the teacher gives back the assignment to stress that writing is an ongoing process. This could reduce the anxiety for the graders and allow for more practice. Other researchers stress the importance of a clear and easily understandable rubric that the students develop themselves to create more buy-in and student engagement. However, the wide

---

<sup>7</sup> Alfaro 69.

<sup>8</sup> Goldfinch Abstract.

<sup>9</sup> Cassel.

<sup>10</sup> Double 485.

<sup>11</sup> Double 483.

variety of rubrics used makes studying the impact of rubrics in peer grading very difficult. Still others recommend putting students together who are at similar academic levels.

### **Peer Grading in AP US History – A Qualitative Description – Part One**

AP classes are also often the most challenging classes offered in high school and can be filled with competitive students. This competition can also lead to some anxiety among students. The College Board provides a relatively clear rubric to award points for essays, which makes peer grading a more plausible activity than perhaps in some other subject areas. In AP Humanities courses, students generally have to write responses to argument-based essay questions that require numerous skills and a high degree of knowledge of the subject matter. These essays are among the most challenging questions on the AP exams and require a great deal of practice. When I graded real essays for the College Board in Lexington KY a few years ago, one of the most common scores I gave was a 0 (out of 7 points!). Many students hadn't even attempted to fulfill parts of the rubric. AP essays may require students to accurately read and synthesize documents to produce a well-structured essay on a required topic in a short amount of time. Or they may ask students to write an essay based solely on their knowledge of the topic. In addition, the essays typically require higher order thinking, which requires them to make thematic connections across different topics and time periods. Students must also learn to support a sufficiently complex argument with relevant evidence and use transition sentences to tie their ideas together in a coherent, organized, and persuasive manner. Complicating this process, some students who enter AP courses have difficulty simply writing clear sentences. Many students struggle to master these skills over the course of the year, especially since on the AP exam itself they must write the essays, in addition to the other types of questions, during a tiring three hour exam.

With these challenges in mind, I set out to give students the opportunity to peer grade in AP US History and AP US Government. My goals included making the grading and writing process more transparent (since students sometimes complain that grading is arbitrary and unclear<sup>12</sup>) and to give students another opportunity to reflect on the skills needed to successfully perform the required tasks on each essay. The hope was that by making essay evaluation clearer, students would feel more comfortable with the process and earn higher scores on the AP test itself. I also hoped students may also begin to feel more comfortable giving and receiving constructive feedback, a skill which would be useful outside the limited realm of academics.

### **AP US History Peer Grading – Part One – Spring 2021 and Fall 2021**

---

<sup>12</sup> Guskey 38.

I used peer grading for the first time with my AP US History class on their document-based essays on the Industrial Revolution. In previous years, I had given small peer grading assignments to my non-AP grade level US History classes, where students graded introductory paragraphs of research papers I had written with the aim of helping them to write their own introductory paragraphs. Students generally enjoyed doing this, and this helped to facilitate class discussions about thesis statements for even quieter students who did not have to worry as much about evaluating their own work or the work of others. Students felt like they could identify the basic weaknesses of introductory paragraphs that had no argument or listed too much detail or irrelevant information. This experience helped to encourage me to do more formal peer assessment.

When I announced that we would be doing peer grading, some of my AP students were a little nervous about showing work to their peers, but I told them I would anonymize the peer grading, which seemed to relax them. However, arranging the students into groups of two and assigning numbers and letters to each student to hide the grader's identity took more time than anticipated. In the end, several students were able to glean whose work they were grading (the class size was 24) anyway, though this did not seem to change the emotional/ social dynamic in the classroom. Students made supportive comments to each other during the grading process when they knew whose paper they were grading. In general, students were familiar with the seven point grading rubric (0-7 points) since we had discussed it extensively earlier in the year and already had written an essay using that rubric. As such, students seemed at ease with the task. The essay students graded was called a DBQ, or document-based question.

Unlike the College Board grading system, I gave the students the opportunity to give .5 point scores for certain areas of the rubric if they were unsure whether to give the point or not. (I give myself this option as well when grading their essays.) Knowing that they could hedge their grading a little may have also mitigated any anxiety students felt. I also shared with them that I use this technique to alleviate my own anxiety about grading accuracy. This type of transparency feels important to communicate to the students. I also want to reward and encourage students for trying to meet the requirements of the rubric even if they don't fully achieve the objectives since it can also be discouraging to see a rubric sheet littered with zeroes.

During the grading process, students were supposed to provide comments on each part of the rubric, though not all students did. Some students did provide helpful and accurate feedback while others did not and only indicated whether the point was earned or not. I did not read the students' entire essays (with the exception of one student who had asked to meet with me to go over her entire essay) since one of my goals in this project was to save time. Instead, I read only the first and last paragraphs of each essay, which contained three out of seven points in the rubric.

Those paragraphs contained both the contextualization and complexity points, as well as the thesis point.<sup>13</sup> These three points were some of the more challenging points. Students regularly missed those points on essays since they required higher order thinking and analysis. I did not evaluate the parts of the rubric that dealt with document analysis. However, despite this, I found the students' evaluation of each other's work in these areas to be accurate. Only a few times did I disagree with the students' decision to give either 0, .5 or the full point for each part of the rubric they evaluated.

Overall, students did better on this essay than they did earlier in the year. The mean score on the peer graded Industrial Revolution essay was 5.5 out of 7, while the mean score on the same type of essay (teacher graded) from earlier in the year was 4.2 out of 7. This improvement might have had several causes. The essay itself might have had an easier prompt and documents. Students could have simply improved their ability to write essays. Some students may have awarded points in the document analysis portions of the rubrics that I did not evaluate and perhaps would not have given. Especially in the relatively tight knit community that SAR High School provides, there is a possibility that students are reluctant to upset their longtime peers. Some students even expressed concern to me that there was nothing wrong with the essays they were grading and as a result did not really know what to do. While most essays were in need of improvement in some areas, it was certainly possible that some students earned all the points given by their peer graders. I simply told students to give them full credit and not worry that they found no particular faults.

I did another round of peer grading the following year. This time, instead of having students grade a document based essay, I had students grade an essay that did not require the use of documents but was instead based on a 6 point rubric, the somewhat misnamed LEQ or Long Essay Question, which is actually shorter than the DBQ. Since the essay was shorter than the document based essay, I graded the entire essay instead of just looking at parts of it. Once again, students' grading accuracy was very strong, with most students giving the same score I gave or a score very close to it. Scores ranged from 3.5-6. Students also gave as accurate and detailed feedback as I gave. This time, I gave students an even more detailed rubric with 2-3 three criteria per point to evaluate, which may have improved grading accuracy. The grading was again anonymous, and students were less likely to determine the identity of their grader than they were last year. I told students that the accuracy of their grading would be taken into consideration when giving their final grade on this project. I did not try to break down how I would do that, however. In truth, awarding points exactly based on grading accuracy or lack of accuracy would have required a lot more work than I was willing to put in. However, just telling the students that their

---

<sup>13</sup> Here, contextualization means that students have to put their argument into historical context. The complexity point refers to the task where students must look at the argument from a different historical lens.

effort mattered and requiring them to write a bullet point or sentence for each part of the rubric, in my opinion, was enough to ensure that a high performing group of students would take the activity seriously, which they did. All students wrote the right number of comments. They did not resort to perfunctory commenting like “good job”, but used more helpful comments, which related directly to the rubric criteria alluding to the level of detail, accuracy, and line of reasoning required for this essay. For example, one student wrote, “explained the concept of market revolution in a clear way which transitioned to the thesis well.” Another student wrote, “not very specific and does not give enough context for the argument - what was it like before these inventions?”

### **Peer Grading in AP US Government Spring 2021**

Similar to AP US History, AP US Government also requires students to write an argument-based essay. The essay has a rubric out of 6 points. We first discussed the rubric in class and then I gave students two practice essays to grade from the College Board, which we spent one full period working on. Students’ comments on the practice essays seemed basically accurate, though they were not asked to provide any written feedback. Instead, I walked around the room and asked students what they thought about the essays. We then had a full class discussion about the strengths and weaknesses of each essay.

Since I had two large sections of AP US Government (24 and 25 students each), each section wrote a different essay, which their peers from the other section then graded. This was the first full essay students had written all year long with the College Board rubric. I assigned the essay towards the end of the year, near the AP Exam, so students would have the rubric in mind when taking the AP exam itself.

The AP Gov students had less practice than the AP US History students. However, the AP US Government essay is easier to write for most students than the AP US History document-based essay since there are no documents and less outside knowledge is needed. In addition to having only 6 points in the rubric, there was also less higher order thinking involved. Arguably, there is really only one part of the rubric, the rebuttal at the end of the end of the essay, that requires some deeper analysis. As a result, this type of essay should be easier to write and grade since students have fewer complex tasks to complete. Some parts of the rubric are also really easy to grade, such as whether students provide evidence simply *relevant* to the prompt.

However, another part of the rubric was also subjective: whether a student sufficiently explained their reasoning throughout the essay. This could present some challenges defining what “sufficient reasoning” entailed. In addition, given that this assignment was done towards the end of senior year after college admissions decisions had been made, I was a little concerned that some

students would not take it as seriously as assignments earlier in the year. On the other hand, this was an AP course, and I had many strong students in both sections, which I hoped would offset that concern.

Unlike some of the earlier AP US History essays, after the students submitted the essays I decided to grade all of them individually in their entirety instead of just parts of them. Despite the fact that it meant grading 49 essays, these essays were relatively quick to grade because they did not have documents associated with them, making them shorter than the AP US History DBQ essays and also shorter than the AP LEQ essays. Yet grading such a large number was still, of course, time consuming and took several hours over spring break.

Overall, most students took the peer grading seriously. Students graded the essays anonymously and, unlike in the APUSH peer grading, there was little chance of knowing whose paper was being graded because each section graded the other section. Thus, the graders were not in the same room as the authors of the essays. I gave them a rubric sheet where they could leave comments and tally the essay scores in a more formal manner. Many students graded more than one essay since the essays were relatively short and could be graded relatively quickly. So some essays had 3 graders: myself and two students. Again, the creation of anonymity with letters and numbers was time consuming. When I gave the essays back, some students were tense because the process of writing and peer grading was more time consuming than normal, but when I explained that the grade itself was not a major assessment the students relaxed.

The overall grades they gave, in my view, were fairly accurate. There did not appear to be a strong bias towards grading more harshly or more leniently. Many students gave the exact same grade that I gave, ranging from 2.5-6. No one scored lower than 2.5 though it was possible to do so. Only a few essays were off by more than one point, and these grades were given by students who were probably not taking the assignment seriously. Some students, in my opinion, actually graded more accurately than I did when I evaluated their comments and scoring decisions. I found myself giving back a point or two in a few places after looking at their feedback. Other students strayed from the rubric and gave points in increments even smaller than the half point rubric, an unhelpful practice. Overall, stronger and weaker students did not vary much in accuracy, as found in other studies. At times, some of the grading errors came from a misconstrual of the essay content itself (for example, not understanding the rulings in certain Supreme Court cases). Other times, some of the decisions that students made were not what I have done but were still defensible. For example, to earn the thesis point, a student had to take a position on the question and establish a line of reasoning. In some cases, student graders and I differed over whether or not a line of reasoning was established, but when I went through those areas of disagreement, their decisions made sense and their commentary on the thesis was generally sound. Students



were also good at identifying whether a part of the rubric was addressed at all but had more difficulty in determining whether a point was fully justified. This also came up most frequently when student graders had to determine whether an essay was argued in enough detail to qualify for the “sufficient reasoning” point. Going forward, defining the word “sufficient” more clearly would be helpful.

The quality of the feedback varied widely. Some student graders simply gave the points without any commentary at all. While that did not mean that they graded inaccurately, sometimes they did. This was fortunately uncommon, and this feedback was given by students who had basically stopped working hard in the class, including one who had earned an A during the year but then went on to earn a 2 on the exam itself, a failing score. A few others wrote some unhelpful comments like “nice work.” Most students, however, attempted to write some meaningful comments. In fact, some students almost gave the exact same feedback I had given; they just used slightly different language. This repetition can make the feedback even more powerful. Overall, I wrote more comments than the average student, but in many cases I felt students gave meaningful feedback. In the end, I did not really hold students accountable for their grading despite giving them a verbal warning to take this activity seriously. This may indicate that I should read through student essays another time to comment on the accuracy of their comments since I too can make errors over the course of grading a large number of essays. I did this in several cases, but not uniformly. However, this may not necessarily be fully worth my time. Perhaps mandating a second grader in all cases would be better. I also did not emphasize that students should write commentary for each part of the rubric, something I probably should have done.

## **Conclusion**

Peer grading can be a useful tool for grading AP essays. Students get more practice with essay rubrics, making the essays seem less intimidating and the scoring less arbitrary. Some students were concerned that peer graders may substantially lower their own essay scores, but in practice this never happened; I always looked at the essays where student graded essays were lower than the scores I gave. To look at each student’s comments and formally evaluate their ability to peer grade can be a somewhat time consuming task, however, so I don’t recommend doing this on a regular basis. Instead of formally evaluating a student’s peer grading, I think simply giving students a detailed rubric and verbal warning to take the activity seriously is sufficient for AP classes. Another time saving approach would be to simply have students grade the two paragraphs of an AP essay and decide whether or not the student earns the points according to the rubric. Breaking up the assignment into smaller blocks would be an especially effective strategy for weaker writers. They can then build up to an entire essay as the year progresses.

Finally, I found simply changing the perspective of the student to be valuable. Having students become evaluators forces students to look at assignments from a different perspective and requires them to engage with material in a unique way. This activity also provides a break from the typical assignments students do. Overall, I found that peer assessment can be modified to meet the needs of an AP class in an effective manner and provided a welcome change of pace from teacher-centered assessment.

### **Works Cited**

Cassel, Sean. "Peer Grading Done Right". *Edutopia*. Accessed July 30, 2021.

De Alfaro, Luca and Shavlovsky, Michael. "Dynamics of Peer Grading: An Empirical Study" International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining (EDM), 9th, Raleigh, NC, Jun 29-Jul 2, 2016: 62-69.

Double, K.S., McGrane, J.A. & Hopfenbeck, T.N. The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educ Psychol Rev* 32, 481–509 (2020). <https://doi.org/10.1007/s10648-019-09510-3>.

Goldfinch and Falchikov. "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks". *Review of Educational Research*. 70, no.3: 287-322.

Guskey, Thomas. *On Your Mark, Challenging the Conventions of Grading and Reporting*. Solution Tree Press, Bloomington, 2015.

Joe Feldman, *Grading for Equity*. Corwin, 2018.

Freeman, Scott and Parks, John. "How Accurate is Student Grading". *Life Sciences Education*, 9, no. 4 (2017): 482-488.

Kohn, Alfie. Who's Cheating Whom? [Alfiekohn.org](http://alfiekohn.org), <https://www.alfiekohn.org/article/whos-cheating/> accessed July 23, 2021.

Sadler, Philip and Good, Eddie. The Impact of Self and Peer Grading on Student Learning. *Educational Assessment*. 11, no. 1: 1-33.

Topping, Keith. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 68, no. 3 (1998): 249-76.