

When “Rigor” Backfired: My Worst Assessment

I remember my worst assessment clearly because it was my first full year of teaching high school social studies. Wanting to bring some of what I enjoyed in my AP classes in high school into my own classroom, I sought to challenge my students. To do this, I developed my own document-based question (DBQ) centered around a complex and controversial prompt: Should the United States have dropped the atomic bombs on Hiroshima and Nagasaki? To help accomplish this I gathered various primary and secondary source documents for students to draw on when crafting their responses. These included excerpts from the Japanese war council meeting before the first bomb was dropped and after, news articles about the bombing and reasons why they were dropped in the aftermath, and excerpts from Truman's diary before and after the bombings. On paper this assessment looked strong. In practice, it failed.

When I had introduced the DBQ, some students began to panic. I had given verbal reminders about the DBQ in the weeks to days leading up to the assessment. However, I made one fatal flaw: I failed to provide meaningful opportunities for students to practice the various skills required. Students were suddenly expected to read sometimes dense documents, analyze perspectives, weigh evidence that I provided, and then construct an argument within one sitting. Many had a hard time figuring out where to start; others attempted responses, but their work showed just how confused they were.

The problem wasn't my content – it was my assessment process. I had skipped essential instructional steps and violated a core principle of backward design (Wiggins & McTighe, 2005): aligning instruction with the desired outcomes. While I had a learning objective for students to evaluate historical decisions using the provided evidence, I had not scaffolded the skills required to reach that objective. There were no smaller checkpoints, no structured practice, and no opportunities for formative assessment in the lead up to the final task. Instead, I bombarded my students with everything at once, which overwhelmed them and obscured their actual learning.

Another issue with this assessment was validity. A valid assessment should measure what it intends to measure. I had intended to assess students' ability to evaluate various historical evidence and develop an argument. Instead what I actually measured was their familiarity with DBQ structure—something I had not explicitly taught. The result was that the assessment did not produce any meaningful or accurate information about my students' understanding.

If I were to redesign this unit today, I would build toward the DBQ gradually so students would not become so overwhelmed with the amount of work. This could be done through having students analyze historical documents individually and in group settings to practice writing claims and supporting them with evidence. I could then incorporate formative assessment strategies through things such as exit tickets and short written responses to help gauge readiness and provide constructive feedback. This way, by the time students reached the final DBQ, they would have both the skills and confidence to succeed.

Ultimately, this “worst assessment” became one of my most valuable lessons, as it helped me make a shift from viewing assessment as a final judgement to understanding that it needs to

be an intentional part of the learning process. Rigor does not mean throwing students into the deep end – it's about teaching them to swim first.

References

Wiggins, G., & McTighe, J. (2005). *Understanding by design*, (2nd Ed.). Association for Supervision and Curriculum Development.