Metropolitan Business Academy: A Case Study in Performance-based Assessment

Ethan Young
Yale College, New Haven Connecticut

Abstract:

This paper examines a New Haven, CT high school where all students take performance-based assessments (PBA) in addition to standardized tests, including the Scholastic Aptitude Test (SAT). Widely cited by scholars as pedagogically sound and logistically unworkable, PBA presents an alternative to standardized tests for evaluating student learning. Beyond potential pedagogical benefits, Bloom’s Taxonomy suggests the promise of PBA to capture complex thought. Using student survey data and teacher interviews, this analysis considers rigor, alignment, and students’ metacognition to explain higher student success rates on performance based tasks than SAT indicators at Metropolitan Business Academy (MBA). MBA students recognized differences between these two assessments, and strongly preferred PBA for its pace, fairness, and opportunities for feedback. Although students found the SAT more difficult to prepare for and take, they felt that PBA more closely correlated with instruction and more accurately reflected their knowledge and skills. These findings suggest the value of authenticity in performance-based tasks to drive student motivation. MBA students viewed performance based assessment as a meaningful exhibition of their learning and this belief motivated higher performance rates than those exhibited during comparable SAT tasks. Further research should investigate these trends in different subject areas (namely mathematics) and among different student populations.

Suggested Citation: Young, E. (2018). Metropolitan Business Academy: A Case Study in Performance-based Assessment (Unpublished Education Studies capstone). Yale University, New Haven, CT.
Yale College

Metropolitan Business Academy: A Case Study in Performance-based Assessment

Ethan Young

Berkeley College

Adviser Professor Richard Hersh

Yale University Education Studies Program, Senior Capstone Project
### Table of Contents

Abstract

Introduction

Scope and Research Questions

Literature Review

Methodology

Findings and Analysis

*Ensuring Rigor: The Moderation Study*

*Alignment with Curriculum and Instruction: The Performance-Based Assessment*

*Recognizing Cognitive Processes: The SAT and PBA*

Limitations

Discussion and Conclusion

Appendix A: Teacher Interview Questions

Appendix B: Student Survey

Appendix C: Moderation Study Rubric

Appendix D: Moderation Study Planning 2017-2018 (abridged)

Acknowledgements

Works Consulted
Abstract

This paper examines a New Haven, CT high school where all students take performance-based assessments (PBA) in addition to standardized tests, including the Scholastic Aptitude Test (SAT). Widely cited by scholars as pedagogically sound and logistically unworkable, PBA presents an alternative to standardized tests for evaluating student learning. Beyond potential pedagogical benefits, Bloom’s Taxonomy suggests the promise of PBA to capture complex thought. Using student survey data and teacher interviews, this analysis considers rigor, alignment, and students’ metacognition to explain higher student success rates on performance-based tasks than SAT indicators at Metropolitan Business Academy (MBA). MBA students recognized differences between these two assessments, and strongly preferred PBA for its pace, fairness, and opportunities for feedback. Although students found the SAT more difficult to prepare for and take, they felt that PBA more closely correlated with instruction and more accurately reflected their knowledge and skills. These findings suggest the value of authenticity in performance-based tasks to drive student motivation. MBA students viewed performance-based assessment as a meaningful exhibition of their learning and this belief motivated higher performance rates than those exhibited during comparable SAT tasks. Further research should investigate these trends in different subject areas (namely mathematics) and among different student populations.
**Introduction**

In seventh grade, I first became involved in education policy by writing the Tennessee Department of Education regarding its mandatory, state-wide assessments. Each year, I sat for three full schooldays bubbling in answers and chewing peppermints meant to stimulate my hippocampus. The Tennessee Comprehensive Assessment Program (TCAP) was a dramatic event: local news media interviewed students, teachers waited anxiously as their score-aligned salaries were determined, and politicians anticipated a new year of results that could improve Tennessee’s national reputation. Suffice it to say that I grew up regularly taking “high-stakes” tests. Despite the fanfare surrounding TCAP, its test questions were often ambiguous, poorly-designed, and reductionist. Moreover, the TCAP “experience” was an intrusion into my learning, not an enrichment of it. I noted these complaints in a four-page letter to Nashville, but never received a reply. My frustration persisted: how could this low-quality exam convey my intellectual capacity or growth?

Since middle school, I discovered educational research that articulated my frustration with standardized tests. In 1956, Benjamin Bloom published his *Taxonomy of Educational Objectives*, which classified cognitive processes and forms of knowledge into hierarchies of complexity (see Figure 1). Since then, Bloom’s Taxonomy has been revised and critiqued, but its central contribution endures: educational tasks require different degrees of cognitive processing, the most complex of which is creation—a new idea, investigation, or product (Anderson et al., 2001). In other words, completing a multiple-choice exam demands different levels of cognition than writing an essay or articulating an original thought. At best, standardized exams ask students to remember, understand, apply, analyze, and occasionally even evaluate; but, they rarely gauge creation, the highest form of cognition according to Bloom’s Taxonomy (Anderson
et al., 2001). So, perhaps my seventh-grade self reacted to the agony of enduring the TCAP for three days, the pressure of the examination environment, or its poorly-crafted materials. Or, perhaps I reacted to a deeper flaw: standardized tests were assessing lower-level cognitive processes, tasks I found dull and uninteresting. My experience in Tennessee’s public schools, however, was hardly unique.

Throughout the past two decades, standardized tests have come to dominate American education. In the pursuit of accountability, legislators have increasingly sought discrete, quantifiable means by which to evaluate schools. To generate data about something as complex and mysterious as learning, assessments necessarily rely on reductionism—most commonly exemplified by multiple choice questions. Meanwhile, educators ask their students to do just the opposite: to broaden one’s mind and formulate complex ideas. Often, both parties end up frustrated. Policymakers call for better data or dramatic reforms based primarily upon test scores. Teachers bemoan the intrusion of testing into their instructional time. Buried beneath this debate sit students, expected to learn and perform on a range of assessments. Most young Americans
take tests just like the TCAP—high-stakes, comprehensive, and authoritative in the minds of many policymakers and educational leaders.

Standardized assessments prevail due to the acceptance of a pedagogical narrative in American culture and politics: these tests enable a social meritocracy by consistently and fairly measuring academic ability. Correlatively, students’ performance on prevalent standardized tests indicates, and often serves as the only legitimate indication of, certain content and skill mastery. This narrative not only permeates education scholarship but also fundamentally shapes our understanding of human worth and social possibility. Although literature supports the validity and reliability of standardized testing in what they purport to measure, dissident voices have proliferated in recent decades—educators, researchers, and policymakers (Darling-Hammond, 1994; Gordon & Rajagopalan, 2016; Hursh, 2005; Koretz, 2017; Ravitch, 2013; Rothstein, 2004; Schneider, 2017) questioning the validity of such testing.

Their critiques vary. Some lament a conflict of interest linked to for-profit testing behemoths driving much of today’s standardized test production and evaluation (Ravitch, 2013). Others point to the significant loss of instructional hours due to frequent in-class testing requirements (Wilczak et al., 2014). Still more critique the way in which test data has become an end of teaching in its own right, surpassing more essential matters like learning and growth (Koretz, 2017). But some have leveled more fundamental critiques: standardized tests are intrinsically unfair and inconsistent; their very design preordains that certain students perform well, and others do not (Darling-Hammond, 2007; Lee, 1998; Rothstein, 2004). According to these critics, performance divisions fall along familiar categories of race, class, and gender. Many also lend credence to my seventh-grade observation. Referencing Bloom’s Taxonomy, they question the efficacy of standardized exams to assess the most complex learning outcomes,
such as creativity and problem solving (Gordon & Rajagopalan, 2016; Koretz, 2017). If true, these criticisms reanimate a time-honored question: how should we go about judging human achievement and potential? Can assessment provide robust accountability while maintaining fidelity to our educational aspirations, such as equal opportunity and transformative learning for all students?

While educators, researchers, and policymakers debate these questions, a movement began in New York City twenty years ago. Thirty-two public schools formed a coalition—the New York Performance Standards Consortium (NYPSC)—to support a common conviction: performance-based assessments (PBA) offer superior insight and learning opportunities to students than those of standardized tests. These performance-based assessments ask students to create products—essays, oral presentations, portfolios—that demonstrate their mastery of standards like argumentation, source evaluation, and contextualization. After developing rigorous testing materials and lobbying the New York Department of Education, NYPSC schools were exempted from New York State Board of Regents exams—New York’s TCAP equivalent—allowing performance-based tests to serve as indicators of student growth and achievement. Moreover, NYPSC schools averaged a ten percent gain in 4- and 6-year graduation rates, as well as a twenty-percent gain in college enrollment 18 months post-graduation (Redefining Assessment, 2017). Their success inspired other educators to develop alternatives to high-stakes testing programs.

In 2013, educators from a Connecticut high school visited a NYPSC school and adapted its methods. Metropolitan Business Academy (MBA), an interdistrict magnet high school located in New Haven’s Wooster Square community, features performance-based assessment as a required component of every student’s education. I first learned about MBA’s unique assessment
program in conversation with Principal Judith Puglisi, who pioneered the initiative. She showed me videos of high school seniors delivering independently-researched presentations to community members, teachers, and peers. These oral presentations, along with extended research papers and round-table discussions, culminate four years of work at MBA. Students choose their subject and presentation format, but everyone must produce a substantial artifact that demonstrates mastery of specified competencies. Although students continue to take some state-mandated standardized tests, MBA assigns significant value to their in-house PBA program for its ability to improve instruction and empower students by equalizing the playing field. Like NYPSC schools, MBA saw broader student improvements after its assessment shift.

Four years into the program, an extraordinary trend emerged: students performed better on PBA than the College Board’s Scholastic Aptitude Test (SAT). All eleventh graders at Metropolitan complete a performance-based Moderation Study and the Evidence-Based Reading and Writing portion of the SAT. These two exams purport to test nearly identical outcomes: reading comprehension skills, evidence-based arguments, and evaluation of sources. In 2016, 37.1 percent of eleventh graders met a College Ready standard as assessed by SAT’s English section.1 Meanwhile, 60.0 percent of those same students were identified as “Competent” or above on Metropolitan’s performance-based Moderation Study.2 Moreover, the Moderation Study is a diagnostic test administered to gauge students’ abilities before they enter an instructional year. Thus, nearly a third of Metropolitan’s eleventh graders were given two verdicts about their preparedness for the future.

1 Data from “New Haven Public Schools, Metropolitan, High School Balanced Progress Report.”
2 From the outset, it is important to clarify the impossibility of direct comparisons between PBA and SAT scores given the two assessments’ distinct methods. I treat this subject in greater detail in the Methodology section.
Why are students succeeding at higher rates on an assessment that is, in theory, more challenging? Three explanations come to mind. Maybe the performance-based test is easier or more lenient, so more students earn higher scores. The performance-based test might better align with students’ experiences in the classroom, giving them a conceptual advantage going into the task. Or, perhaps students view these two tasks very differently, and apply themselves more fully towards the performance-based task than the SAT. These explanations yield research questions about rigor, alignment, and student metacognition.3 Are Metropolitan Business Academy’s performance-based assessments rigorous, or can these data be ascribed to lower standards? Are students benefiting from clearer correspondence among curriculum, instruction, and assessment, or do performance-based assessments and the SAT align similarly? Are students aware of the cognitive processes being probed in performance-based assessments as compared to SAT, or do they understand both assessments to be similar? By addressing these three questions, I will construct an explanation for these unexpected data.

Scope and Research Questions

Assessment is among the most widely debated, heavily researched, and politically charged topics in education. Consequently, it is important to address this study’s narrow scope from the outset. My project focuses on one performance-based assessment administered to all MBA students. Regardless of Academic Pathway, everyone completes a writing-based PBA during their junior year.4 My research evaluates the rigor of Metropolitan’s performance-based

---

3 These terms are further defined in the Literature Review section, but to offer preliminary definitions: rigor refers to the fidelity of an exam’s results to the purported objectives being tested; alignment refers to the correspondence among curriculum, teaching, and assessment; student metacognition refers to the ability of students to recognize their own cognitive processes.

4 MBA students choose an Academic Pathway, options of which include Law and Political Science, Allied Health and Science, Digital Arts and Technology, and Finance.
assessment, contextualize program alignment in existing scholarship, and evaluate students’ metacognitive distinctions between performance-based assessments and standardized tests, namely the SAT. With this information, I attempt to make sense of the finding that more students perform well on performance-based assessments. In doing so, I contextualize Metropolitan’s program design in broader research, and offer a case study of performance-based assessment to the larger field of Education Studies.
Literature Review

This analysis intersects scholarship concerning assessment theory, educational psychology, and the history of education policy. Before describing specific assessments, one should define the term. An assessment, or test, is a formal attempt to determine a student’s relationship to specified variables (i.e., learning outcomes, attitudes). As Popham (2003) outlines, educators rely on assessment to enable decisions about “the nature and purpose of curriculum,” “students’ prior knowledge,” “how long to teach something,” and the “effectiveness of instruction” (p. 9). Thus, assessment lies at the core of educational practice, facilitating exchange between curriculum and instruction (Anderson et al., 2001; Popham, 2003). Alignment describes the extent to which these three elements correspond and interact (Anderson et al., 2001). As the cohesion between instruction and evaluation, curriculum and instruction, or evaluation and curriculum improve, so too does alignment.

Assessments range considerably, but they can be classified by their role in an educational system and the kind of learning they evaluate. Most assessments are diagnostic, formative, or summative (Luongo-Orlando, 2003). Diagnostic assessments, such as MBA’s Moderation Study, provide educators with baseline information about students’ knowledge and skills to create effective teaching plans. Formative assessments, such as MBA’s performance-based assessment, function within the learning process throughout a school year, offering iterative feedback that allows students and teachers to adjust their methods to meet performance goals. Summative assessments, such as the SAT, occur after a period of teaching and ask students to demonstrate their knowledge and skills on-demand. Diagnostic, formative, and summative assessments achieve different goals, and it is important to distinguish performance-based assessments and the SAT since, “formative assessment is used primarily to improve
student learning; summative assessment is used primarily to assign grades” (Anderson, et al., 2001, p. 102). So, how do these different assessments evaluate student learning?

**3.1 The Taxonomy Table**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Factual Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Conceptual Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Procedural Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Metacognitive Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Educators design assessments to evaluate certain cognitive processes. Anderson, et al. (2001) revised Bloom’s Taxonomy (Bloom, 1956) and present a Taxonomy Table for classifying assessments in terms of cognitive processes and types of knowledge. The Cognitive Process Dimension describes the form of cognitive work required by a student in a given task, ranging from memorization on the least difficult end to creation on the most difficult. The Knowledge Dimension describes the type of knowledge required by a task, ranging from factual knowledge on the least difficult end to metacognitive on the most difficult. As visualized in Figure 2, each tier in in the taxonomy is thought to be more difficult and, by extension, less

---

5 There are two essential merits to describing assessments with the Taxonomy Table. First, the Table enables detailed analysis of alignment by displaying the correspondence among curriculum objectives, how they were taught, and assessment of student work. In other words, it reveals the distance between theory and practice in education. Second, the Table assists educators in eliciting high-level cognitive processes among students.
commonly achieved by students. Kreitzer and Madaus (1994) support the hierarchical theory of Bloom’s Taxonomy, which suggests that students progress through each step of cognitive processing sequentially. This hierarchy further suggests that only by considering the long-view of a given educational task is it possible to guide students to creation and evaluation. Using the Taxonomy Table, comparisons of performance-based assessments and the SAT reveal underlying differences in their educational philosophies and abilities to test cognitive processes.6

Performance-based assessment (sometimes called ‘performance assessment’) has a longstanding and robust presence in educational research and pedagogical literature (Rudner & Boston, 1994; Pecheone et al., 2010; Darling-Hammond, 1994; Baker et al., 1993). PBA’s underlying philosophy appeared as early as John Dewey, who wrote: “The child’s own instincts and powers furnish the material and give the starting-point” (Dewey, 1964, p. 428). Although performance-based assessments range considerably in format and scope, they consistently entail student-produced artifacts (i.e., portfolio, oral presentation, research paper) and timely feedback from educators (Rudner & Boston, 1994). In the United States, performance-based assessments rose to national awareness in the 1980s as indicated by the U.S. Department of Education’s review of performance-based assessment in its Educational Resource Information Center (ERIC) publication (Rudner & Boston, 1994). Research literature offers numerous definitions of performance-based assessment, but most studies share key traits.

Performance-based tests are defined by their authenticity, role in the learning process, and capacity for individualized learning. Luongo-Orlando (2003) distinguishes performance-

6 It is important to acknowledge the perspective of the Taxonomy Table’s creators. Anderson, et al. (2001) note the potential for performance-based assessments to measure high-level cognitive process, but also acknowledge the additional burdens of time, test scoring, and maintenance of student records.
based tasks by their authenticity, meaning they “require the application of knowledge and skills in real-life situations” (p. 7). Svinicki & Achacoso (2005) emphasize the use of project-based learning as a means for establishing these real-life situations, requiring incremental, self-directed work. Educator feedback and rubrics also underlie any performance-based task. Brualdi (1998) underscores the importance of selecting clear criteria for assessment and creating corresponding rubrics based on the curriculum. Darling-Hammond (1995) highlights educators who record descriptive accounts of student progress throughout the year, which could be discussed with other educators about a student’s multi-year intellectual journey. Lastly, “individually configured excellence” centrally shapes the performance-based learning environment (Gardner, 1991) by enabling “assessment practices [that] are able to support high standards without standardization—to promote excellence in the context of diversity” (Darling-Hammond, 1995).

In other words, these assessments enable students with diverse backgrounds, experiences, and learning approaches to produce meaningful reflections of their learning in an individualized manner. Whatever its format and evaluation, PBA has received substantial praise from scholars.

Several key arguments for performance-based assessment have emerged: they improve alignment among instruction, evaluation, and curriculum (Darling-Hammond, 1994); they provide more nuanced information about student progress over time (Janesick, 2006); and they feature prominently within successful education systems in Finland, Sweden, Australia, Hong Kong, Singapore, and the United Kingdom, among others (Darling-Hammond, 2010). Linn et al. (1991) argue that standardized tests are “indirect indicators” of students’ underlying abilities and judge performance-based assessments to be “direct” measures of ability. Janesick (2006) notes how timely feedback provides students with opportunities for growth. Most importantly, scholars agree that performance-based tasks require higher-level cognition (Anderson, et al., 2001). This
improved evaluation of student cognition offers new opportunities for both students who
traditionally underperform on standardized tests and for exceptionally-talented students (Van
Tassel-Baska, 2013).

Although scholars generally support performance-based assessment on a theoretical
basis, many dispute its practicality. Linn et al. (1991) question whether and how performance-
based assessments can be evaluated for fairness, cognitive complexity, content quality,
generalizability, cost, and efficiency. In other words, PBA comes closer to student learning but
potentially at an insurmountable logistical cost. Wiggins (1990) resolves this tension by
advocating for a more labor-intensive form of assessment since standardized tests give educators
a false impression of academic growth. Similarly, Burger & Burger (1994) highlight issues of
reliability in performance measures, noting that standardized testing has gained respect from
teachers for its reliance on norm-referencing (a measure of whether test takers perform better or
worse than a hypothetical “average” student); however, they emphasize that subjective judgment
in authentic assessments are unavoidable and that these measures are necessary to measure real
learning. Woolley (1996) echoes this sentiment, disputing the theory of score validity in
standardized testing. While these researchers ultimately accept the risk of less consistent
evaluation, others disagree.

Some researchers approach performance-based assessment with greater hesitation. Baker
et al. (1993) suggest that educators and assessment developers lack experience balancing breadth
and depth to implement alternative assessments effectively. Haertel (1990) and Hood (1998)
reiterate concerns about accountability, and Hood notes the potential of performance-based
assessments to allow for some educators’ racial biases to negatively impact students of color.
These implementation concerns can be reduced to primarily administrative oversight and costs
Metropolitan Business Academy: A Case Study in Performance-based Assessment

(Schneider, 2017). However, Schneider (2017) notes the promise of technology in reducing logistical burdens of performance-based assessments, such as computer simulations that assist in scoring results. Schneider (2017) further notes that districts could reallocate their existing professional development expenditures towards performance-based discussions that would both facilitate faculty dialogue and enable assessment implementation. Kamenetz (2015) echoes Schneider, suggesting that widely-implemented performance-based assessments may cost districts less overall, citing the Performance Assessment Review Board, Inc. that serves thirty-nine New York schools with an annual budget below $0.5 million. Regardless, researchers approach performance-based assessment cautiously and skeptically. However, their hesitation has not prevented numerous states from systematizing PBA.

American states have frequently experimented with performance-based assessment in their education policies. Pecheone et al. (2010) highlight states that have already incorporated performance-based assessments within large-scale accountability systems, including Connecticut, New York, and New Jersey. In particular, they underscore Connecticut’s long-standing commitment to performance assessment as manifest in its Connecticut Academic Performance Test (CAPT), which measures student progress towards broad educational goals across academic disciplines (Pecheone et al., 2010). Furthermore, the authors reference a longitudinal study that correlates high CAPT scores among test-takers to post-graduation success in college and career (Pecheone et al., 2010). Within their evaluation of ten state case studies, Wilczak et al. (2014) find that Connecticut’s Mastery Test (now retired) embodied best research

---

7 While Northeastern states led the nation in widespread policy implementation of performance-based assessment, other states have implemented smaller programs. For instance, South Carolina offers Performance Task Assessments (PTA) for students identified as talented and gifted in grades 2-5.
practices surrounding performance-based assessment and enjoyed long-standing success thanks to high levels of community involvement and support. These programs largely came and went, however, and were ultimately sidelined by standardized measures. Darling-Hammond (1994) addresses these fleeting policy experiments by warning against their widespread implementation of PBA. Instead, she argues for “top-down support for bottom-up reform,” in which schools are empowered to create formative and useful assessments (Darling-Hammond, 1994).

Performance-based assessment has shaped numerous international education systems. Drawing upon the educational practices from high-performing nations, Darling-Hammond & Wentworth (2010) contend that performance assessment plays an effective and natural role in bridging curriculum, instruction, and assessment in Finland, Sweden, Australia, Hong Kong, Singapore, and the United Kingdom. For instance, Singapore relies heavily upon project-based work and oral presentations (Darling-Hammond & Wentworth, 2010). Two articles from Blömeke & Gustafsson (2017) investigate the role of PBA internationally: Skar et al. (2017) praises low-stakes writing assessments employed in local Norwegian settings and Vainikainen et al. (2017) track the decentralization of Finnish education since the 1990s, a shift that afforded the kind of local implementation espoused by Darling-Hammond (1994). In the U.K., Simpson (2006) points to performance-based assessment and authentic assessment as robust alternatives to current testing regimes in Scotland. The Netherlands and Singapore shifted their schools to emphasize fewer standardized assessments in recent years (Koretz, 2017). Thus, a diverse group of high-performing nations have implemented performance-based assessment, recognizing its benefits for students and for educational leaders.
Turning now to standardized assessments, the SAT has become the most pervasive test for evaluating high school students’ readiness for college and career.\(^8\) After a recent overhaul, the test-makers claimed: “Assessment must do more than simply provide a score or a ranking; it must become a force that helps deliver opportunity. To reach its full potential, the SAT must grow beyond what it has long been—a valid and reliable measure of what research tells us is necessary for college and career readiness and success” (The College Board, 2015, p. 13). Indeed, research lends credence to the SAT’s ability to predict long-term student outcomes. Furthermore, Vogler (2002) finds that the presence of high-stakes performance assessments in Massachusetts public schools caused teachers to change their instructional practices and that those changes were more aligned with what research suggests are “best practices.” From a community perspective, Phelps (2005) finds that American parents, more so than administrators or policymakers, value standardized tests as a means of reflecting high academic achievement.

There are numerous critiques of the SAT, however, that dispute its efficacy in identifying current or future student achievement. Kamenetz (2015) summarizes the critiques well: “testing the wrong things,” “wast[ing] time and money,” “making students hate school,” “making teachers hate teaching,” and “teaching to the test,” among others (p. 13). Kamenetz (2015) cites a recent MIT study that found little improvement in eighth graders’ intelligence despite improvements in standardized test scores. Gordon & Rajagopalan (2016) published the results of the Gordon Commission on the Future of Assessment in Education, which found that standardized tests narrowly evaluate cognitive processes, neglect forms of student diversity,

---

\(^8\) It is important to note that while I focus on the SAT for this study, several standardized exams exist that are considered rigorous, thoughtful, and meaningful for analysis. Included among these are the Programme for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP).
overemphasize comparative standings among students, and privilege knowledge rather than thought. Instead, the Gordon Commission advocates mastery-based assessment in which mental processes are assisted and evaluated rather than demonstrations of knowledge retention (Gordon & Rajagopalan, 2016).

In sum, scholarly literature widely supports performance-based assessment though it remains unclear where and how the method should be employed in schools. In evaluating MBA, this literature plays an important role by both substantiating many aspects of its present implementation while underscoring the difficulty of scale and objectivity. My project will consider MBA as a case study of the practices researched here: to what extent are this high school and the literature congruous?
Methodology

This study occurred on-campus at Metropolitan Business Academy located at 115 Water Street, New Haven, CT from January 2018 through May 2018. A written survey was distributed in March to all current MBA seniors who, by default, completed a social studies performance-based assessment during the previous school year. I also interviewed several MBA teachers. Finally, I considered school-wide SAT scoring trends.

The written survey consisted of two pages and ascertained student perceptions regarding assessment. It was designed for 12th-grade students who are aged approximately seventeen to nineteen years. MBA’s student population is diverse, with students of African American, Hispanic, Asian, Native American, and White ethnicities. All access was facilitated through the Head Principal of Metropolitan Business Academy, Judith Puglisi. The survey contained a mixture of open-ended and multiple-choice questions, which rendered quantitative and qualitative datasets for analysis. The open-ended questions invited student opinion about performance-based assessments, the SAT, and the interaction between classwork and assessment. Several multiple-choice questions asked students to indicate their engagement with various components of PBA (research, presentation, collaboration, etc.) and to compare their engagement between PBA and the SAT. Engagement was measured on a sliding numerical scale.

---

9 The Yale University Human Subjects Committee determined that this research presents minimal risk to subjects. The Institutional Review Board found this study to meet the requirements of 45 CFR § 46.404 in that it presents no more than minimal risk to its minor subjects. Permission of one parent or guardian was sufficient to carry out the study. All student survey participants received written consent from a parent or guardian.

10 See Appendix B for a reprinted copy of the survey.
Once collected, I transcribed open-ended student responses into question-specific documents. Then, I coded, condensed, and interpreted the responses to each question.\textsuperscript{11} Common threads—terminological and pedagogical—emerged from these surveys. Multiple-choice questions, on the other hand, were copied into a spreadsheet for graphic and numerical analysis. Running even simple statistics, such as mean and standard deviation, provided further insight into student consensus. Although student-produced data provides insight into the dynamics of PBA, clear biases and analytic limits applied. For example, it was not possible to conclude that PBA is a superior or more accurate assessment than SAT when student responses tend to favor PBA. Instead, this finding reveals deeper metacognitive awareness of performance-based assessment’s aims, which helps to explain differences in student performance.

In addition to the student survey, I interviewed several teachers who were identified by Principal Puglisi. Given the comparisons with the writing and reading sections of the SAT, Principal Puglisi primarily introduced me to educators in the Social Studies Department. These interviews strictly pertained to teachers’ professional perspectives about PBA and were transcribed anonymously for analysis. In some cases, teachers shared anonymized examples of student work with me to further explain aspects of MBA’s approach. Questions were intended to elicit professional opinions regarding the efficacy or value of PBA in classroom use.\textsuperscript{12} In particular, I discussed feedback and the way in which teachers use PBA as a means of targeted instructional improvement.

\textsuperscript{11} My qualitative research methodology was largely derived from Brinkmann, S., & Kvale, S. (2015). \textit{InterViews: Learning the Craft of Qualitative Research Interviewing}. London: Sage Publications.

\textsuperscript{12} See Appendix A for a reprinted copy of interview questions.
Lastly, I referenced school-wide SAT score data. While individual students’ SAT results were deemed too sensitive for research purposes, I obtained generalized data about MBA student performance. It is important to clarify the terms of comparison between the SAT’s Reading and Writing test portion and MBA’s junior year social studies performance-based assessment. The SAT Reading section “emphasizes source analysis and evidence use” (The College Board, 2015). The Writing section asks students to “develop, support, and refine claims and ideas in multi-paragraph passage [and]…analyze a provided source text to determine how the author builds an argument to persuade an audience through the use of evidence, reasoning, and/or stylistic and persuasive devices” (The College Board, 2015). PBA’s Moderation Study evaluates students along three similar criteria: “Form an arguable claim/thesis, support and develop claim/thesis

<table>
<thead>
<tr>
<th>Category</th>
<th>Current SAT</th>
<th>Redesigned SAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Testing Time</td>
<td>3 hours and 45 minutes</td>
<td>3 hours (plus 50 minutes for the Essay (optional))</td>
</tr>
<tr>
<td>Components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Critical Reading</td>
<td></td>
<td>a) Evidence-Based Reading and Writing</td>
</tr>
<tr>
<td>b) Writing</td>
<td></td>
<td>» Reading Test</td>
</tr>
<tr>
<td>c) Mathematics</td>
<td></td>
<td>» Writing and Language Test</td>
</tr>
<tr>
<td>d) Essay</td>
<td></td>
<td>b) Math</td>
</tr>
<tr>
<td></td>
<td></td>
<td>» Essay (optional)</td>
</tr>
<tr>
<td>Important Features</td>
<td></td>
<td></td>
</tr>
<tr>
<td>» Emphasis on general reasoning skills</td>
<td>» Continued emphasis on reasoning alongside a clearer, stronger focus on the knowledge, skills, and understandings most important for college and career readiness and success</td>
<td></td>
</tr>
<tr>
<td>» Emphasis on vocabulary, often in limited contexts</td>
<td>» Greater emphasis on the meaning of words in extended contexts and on how word choice shapes meaning, tone, and impact</td>
<td></td>
</tr>
<tr>
<td>» Complex scoring (a point for a correct answer and a deduction for an incorrect answer; blank responses have no impact on scores)</td>
<td>» Rights-only scoring (a point for a correct answer but no deduction for an incorrect answer; blank responses have no impact on scores)</td>
<td></td>
</tr>
<tr>
<td>Essay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>» Required and given at the beginning of the section</td>
<td>» Optional and given at the end of the SAT; postsecondary institutions determine whether they will require the Essay for admission</td>
<td></td>
</tr>
<tr>
<td>» 25 minutes to write the essay</td>
<td>» 50 minutes to write the essay</td>
<td></td>
</tr>
<tr>
<td>» Tests writing skill; students take a position on a presented issue</td>
<td>» Tests reading, analysis, and writing skills; students produce a written analysis of a provided source text</td>
<td></td>
</tr>
<tr>
<td>Score Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>» Scale ranging from 600 to 2400</td>
<td>» Scale ranging from 400 to 1600</td>
<td></td>
</tr>
<tr>
<td>» Scale ranging from 200 to 800 for Critical Reading; 200 to 800 for Mathematics; 200 to 800 for Writing</td>
<td>» Scale ranging from 200 to 800 for Evidence-Based Reading and Writing; 200 to 800 for Math; 2 to 8 on each of three dimensions for Essay’</td>
<td></td>
</tr>
<tr>
<td>» Essay results scaled to multiple-choice Writing</td>
<td>» Essay results reported separately</td>
<td></td>
</tr>
<tr>
<td>Subscore Reporting</td>
<td>None</td>
<td>Subscores for every test, providing added insight for students, parents, admission officers, educators, and counselors</td>
</tr>
</tbody>
</table>

Figure 2. Comparison of the Major Features of the Current SAT and the Redesigned SAT (The College Board, 2015, p. 18). In 2015, the SAT was redesigned. Here, the parallels with MBA’s assessment are made clear.
using relevant evidence, and analyze evidence” (Moderation Study Rubric). The purposes of these two exam sections are extremely similar—they differ primarily in their methods. Thus, while direct comparison is unproductive, stark contrasts in student performance on these two exams provokes deeper questioning of these two assessments.

---

13 See Appendix C for a reprinted copy of the Moderation Study Rubric.
Findings and Analysis

Ensuring Rigor: The Moderation Study

MBA’s self-designed Moderation Study evidences the rigor of MBA’s performance-based assessments. The study enables educators to establish students’ baseline proficiency, evaluate student performance against common standards, and adjust scoring, instruction, and long-term planning in response. Completed during the first two weeks of class, this initial assessment condenses the yearlong PBA into two weeks by introducing students to a topic and asking them to produce a short research essay. For the 2017-18 school year, teachers proposed a provocative debate questions for essayists: “Should Connecticut follow in New York’s footsteps and make state college tuition free?” (Moderation Study Planning 2017-2018).14 After discussing political cartoons, infographics, and the like with their peers, students took a stance on the issue of college costs and wrote a research paper.15

A common rubric deployed for all four grades allows educators to adjust—or moderate—their instruction and scoring to reflect school-wide performance. In social studies, for example, educators developed a rubric that applies uniformly to all grade levels and considers three components of writing: introduction and thesis, evidence for claims, and analysis of evidence.16 To see the rubric in action, consider two 12th grade thesis statements. Student A wrote, “While Connecticut shouldn’t do exactly what New York did, making college completely free, we

---

15 Teachers select potential articles for use, but also encourage students to search for their own materials. In one document, teachers collaborated to share links of material that assists both sides of the argument, engages at varying degrees of difficulty/reading level, and relies upon different mediums (video, cartoons, columns, etc.).
16 See Appendix C: Moderation Study Rubric. The final rubric employed for the end-of-year PBA is derived from the New York Performance Standards Consortium, but this one is tailored specifically to the needs of MBA.
should do something to make the price of higher education a lot more reasonable and fair.”

Student B wrote, “All I know is nothing is ever free I also know there would be no exception for college are there anyways we can make a compromise?” Student A received an “Exemplary” designation in “Form an arguable claim/thesis within an introductory paragraph” while Student B received “Novice.” This seems apt, as Student A contextualized the issue, took a clear stance, and proposed a nuanced solution. Contrastingly, Student B failed to offer the reader a clear position and drifted into the topic without specificity. Beyond assessing the relative standing of students in a given class year, however, the Moderation Study is intended to track student learning over a four-year progression.

The Moderation Study enables direct comparison of student progress across grade levels by relying on a consistent rubric and assigning the same essay topic. Consider two responses, one from the 9th and one from the 11th grade. The 9th grader concluded, “Free college is a benefit because it will help all kinds of people and families…This can impact society by giving people more jobs.” The 11th grader concluded with, “I believe that college tuition should be free because there will be more successful poor students that wouldn’t have to allow their income define their future.” Both essays received identical scores, with designations of “Competent” in all rubric categories. The two responses share much in common, but the 11th grader does demonstrate a more specific, income-based claim about college tuition’s impact. As one teacher described, these grade-level comparisons facilitate meaningful dialogue among teachers: where do students consistently struggle? Was this a fair task? What kind of progress do we expect each year?

The shared evaluation process among MBA educators illustrates the value of the Moderation Study. After discussion and evaluation, teachers consider school-wide performance within their subject area. The most recent aggregate data from the Social Studies Moderation
Study demonstrates the ways in which these initial PBAs differentiate students, using terms of “Novice,” “Emerging,” “Competent,” “Co-exemplary,” and “Exemplary” instead of conventional letter grades or numerical scores. As Figure 4 demonstrates, students generally excel at higher rates by 12th grade as evaluated by the baseline rubric: the percentage of Exemplary students soars to 23% and Co-Exemplary approaches 20%. At the same time, however, 38% of 12th graders are classified as “Emerging,” a percentage outweighing both the 10th and 11th grade cohorts. While this raises concerns about students’ learning progression over four years, it testifies to the rigor and accountability of MBA’s Moderation Study. Even after three years of working with a particular cohort of students, teachers still evaluate work fairly against the performance of the entire student body.

In addition to matters of relative scoring, MBA educators draw upon the Moderation Study to consider the content of performance-based tasks each year. In a team meeting with the Social Studies department, educators described how their views on students’ topic choice evolved with years of experience. 9th and 10th grade teachers found that their students were wont to choose topics that far surpassed the scope of the assessment: “infinite choice was no choice at all,” said one teacher. The multitude of topic options also limited the amount and quality of feedback teachers were able to provide. For these reasons, educators narrowed the scope of topic selection in 9th and 10th grade by providing a set of themes from which students could select topics. One teacher also noted MBA’s recent subscription to an online database that allowed educators to frame research around pro and con arguments.

Educators also revised aspects of the research process in light of these initial Moderation Studies. As one teacher shared, “We used to require an annotated bibliography…but [it] wasn’t good proof that they were reading. So, [we] changed them to source studies…a printed article,
with annotations, and the corresponding notes.” To further encourage students to engage with their research reading, educators introduced partner discussion groups where, as one teacher described, students ask their peers, “Did you have to look up words? Why is it a good source? It’s helping them to strategize with each other about ways to find good sources.”

Throughout this process of revision, MBA educators used the Moderation Study to redesign performance-based tasks around the abilities and needs of students. This dynamism improves rigor by ensuring that assessments test what they are purported to test each year and students have fair opportunities to excel.

Figure 4. Social Studies Baseline Study, Composite Scores, December 2016, Metropolitan Business Academy
Alignment with Curriculum and Instruction: The Performance-Based Assessment

Educators design assessments to align closely with instruction and curriculum at MBA, and students seem to benefit from clearer alignment. As one teacher underscored, designing classwork and assessment around alignment entails individual sacrifice and consideration of student growth on the whole: “What are you willing to give up with the goal of students being able to prove that they have the skills and habits of mind of a historian at the end of the year? [It] is hard because we love the things we teach, and we invest tremendous time and energy into them…But…if it doesn’t feed in to that final project in some way, then why are you doing it?”

Another teacher explained MBA’s shared design goal: “We really try to plan more purposefully about improving our teaching around class discussion in our instructional design with the outcome of strengthening arguments [and] improving analysis.” This vision of connecting more aspects of class time to performance-based tasks translated into students’ clearer understanding of expectations and engagement with the tasks.

In describing the most enjoyable aspects of PBA, students indicated the extent to which alignment affected their motivation in completing performance-based tasks. Students overwhelmingly indicated that class time and PBA were closely connected, writing “it went along with what we’re learning about,” and “everything we do is connected to each other and is very useful.” While answering the same survey question, many students related this interconnectivity with their overall interest in completing the performance-based tasks. One student described the PBA as an opportunity to “complete a college level paper as a junior,” while another shared that the PBA “allowed me to showcase my learning and write a research paper.” Although these connections were engaging for some students, a vast majority also found the interwoven nature of performance-based assessment challenging.
For MBA students, the most difficult process involved in a PBA was research (see Figure 5). Students were generally less verbose when writing about the aspects of PBAT they disliked, but many mentioned the length and rigor of the process.\textsuperscript{17} “The amount of pages that we had to come up with,” “the amount of topics,” “the long process to complete it,” and “the amount of research” challenged MBA students. One student elaborated on the difficulty of arguing a new position, in particular: “It’s not easy for me formulate an opinion; I often have to go through multiple topics and try them out before I can choose one.” Students also struggled with the conceptual work of research, like outlining “because at times it made writing the paper confusing and frustrating.”

Despite the difficulty of research, MBA students most frequently mentioned the knowledge acquired through the research process as the most enjoyable aspect of the PBAT. This reference to learning outcomes testifies to the meaningful connections between in-class exercises and the final performance-based assessment. Students fondly discussed the content knowledge they acquired through research and even shared their topics in the survey (i.e., the Young Lords or the Black Panthers). Many students chose topics with which they had some familiarity, but one student shared, “it took all my skills and power to...research a topic I knew nothing about.” Some students also enjoyed the synthesis required in research: “being able to combine...my previous knowledge...[and] others point of view on the topic to make one huge statement on it.” These remarks reveal an assessment that motivated students to acquire and understand new knowledge along the process of demonstrating mastery of certain skills, underscoring the power of an assessment tightly aligned to class work and teaching.

\textsuperscript{17} The acronym “PBAT” refers to a specific performance-based assessment test and is largely interchangeable with “PBA.” Four respondents even said they disliked “nothing” about the PBA process.
Students advocated for performance-based assessment in their survey responses, touting its dimensions of alignment, authenticity, and rigor. One student wrote, “PBAT shows more of what you’re capable of doing when given a task,” and another agreed, writing, “I’m more than a grade, number, or rate, and PBAT shows that.” Students also spoke to the rigor of PBATs, sharing that “PBAT was a rigorous process but the end result is top quality,” and “[PBAT] is a better way to measure a student’s knowledge.”

**Figure 5.** Survey Results, “Student Views on Assessment at MBA,” Questions 1-3. For questions 1-5, students were asked to indicate, on a scale from 1 to 5, which option best matches their experience with the given statement (where 1 is ‘Very easy’ and 5 is ‘very difficult’).

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: How difficult was the research process involved in your junior year PBAT? N=42</td>
<td>2.93</td>
<td>0.75</td>
</tr>
<tr>
<td>Q2: How difficult was the presentation involved in your junior year PBAT? N=35</td>
<td>2.31</td>
<td>0.79</td>
</tr>
<tr>
<td>Q3: How difficult was the peer collaboration involved in junior year PBAT? N=39</td>
<td>2.18</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Recognizing Cognitive Processes: The SAT and PBA

Students keenly recognized the different cognitive processes evaluated by each assessment and were more motivated by the inherent rewards of completing a performance-based task. One wrote, “PBAT is more a project that exists in itself, whereas the SAT is a culmination of all the prior years of learning.” Others pointed to the unpredictability of test-taking environments: “I had no idea what the questions would be [on the SAT] which was more stressful. The PBAT was an essay…on a topic I chose.” Another student succinctly captured a key distinction between the two exams: “the SAT covers more broad areas of knowledge, but it better shows what I don’t know than what I actually do know whereas the PBAT is more narrowly focused to one area of knowledge and focuses on what I do know.” By further investigating the kinds of distinctions students drew between these two assessments, it will be possible to discuss how metacognitive awareness affected students’ performance on SAT as compared to the PBAT.

When comparing the SAT and PBAT, MBA students made two clear distinctions. The SAT was perceived to be more “difficult” than any aspect of the PBAT (research, presentation, and peer collaboration), but the PBAT was perceived to more accurately reflect students’ knowledge and skills than the SAT (See Figures 6, 7). When asked which of the two assessments they would prefer to take, an overwhelming majority chose PBATs. Students reasoning ranged widely. Some chose PBAT purely for its ease in comparison to the SAT: “I’m good at writing papers and its easier.” Others pointed to pedagogical reasons, including feedback (“I could…receive help and guidance”), authenticity (“PBAT truly demonstrates my knowledge and skills”), and associated stress (“PBAT…[is] less stressful”). Many students preferred PBA since
it afforded them a greater sense of ownership and accomplishment: “I feel much more proud and accomplished.”

Another key point of comparison among MBA students was the amount of time given for each test’s administration. As one student described, “the SAT…measure[s] my performance on one particular day, while PBAT’s measure it over time.” Another spoke the potential for studying beforehand, “the SAT isn’t something you can really prepare for or demonstrate your skills on. The PBAT requires real skills you will need in college.” This emphasis on flexibility resonated with one interviewed teacher, who said, “there’s no punishment for being behind the timeline. Even if they’re not done, they haven’t failed. You must keep working on it…Let’s roll up our sleeves…I’ve never had a student complain about their grade because if you don’t like your grade then do more work.”

By contrast, many students held the SAT in low esteem by comparison, noting its rigidity and narrow range of tested abilities. One wrote, “I didn’t enjoy sitting for that long during the SAT. PBAT lets our minds run free and lets us express ourselves through projects.” Another said, “PBAT allows you to display your knowledge and creativity through your written words while the SAT is solely based upon a 25% chance.” Many students also questioned the connection between the SAT and their experiences in class. One student wrote, “very few of the questions on the SAT seemed connected to work I did earlier that school year.” Others saw success on the SAT as a function of test preparation alone, “It proves peoples’ testing abilities not knowledge and understanding.” For some students, this test preparation was especially difficult: "As a person who needs to work and do school work, time to study for the SAT isn’t there.”
Figure 6. Survey Results, “Student Views on Assessment at MBA,” Questions 4-5. For questions 1-5, students were asked to indicate, on a scale from 1 to 5, which option best matches their experience with the given statement (where 1 is ‘Very easy’ and 5 is ‘very difficult’).

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4: How difficult was studying for the SAT? N=48</td>
<td>3.58</td>
<td>0.96</td>
</tr>
<tr>
<td>Q5: How difficult was taking the SAT? N=48</td>
<td>3.92</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Figure 7. Survey Results, “Student Views on Assessment at MBA,” Questions 6-8. For questions 6-8, students were asked to indicate, on a scale from 1 to 5, how strongly they agree with the given statement (where 1 is ‘Strongly disagree’ and 5 is ‘strongly agree’).

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6: My junior year PBAT scores accurately reflect my knowledge and skills. N=40</td>
<td>3.75</td>
<td>1.1</td>
</tr>
<tr>
<td>Q7: My SAT score accurately reflects my knowledge and skills. N=47</td>
<td>2.32</td>
<td>1.0</td>
</tr>
<tr>
<td>Q8: The SAT and PBAT are basically the same. N=42</td>
<td>1.76</td>
<td>0.66</td>
</tr>
</tbody>
</table>
Limitations

This project has important limitations. To reiterate, it is impossible to establish a direct comparison between student performance on PBAs and SATs, given that these PBAs are unique to MBA, scored on different numerical scales, and qualitatively, rather than quantitatively, assessed. Furthermore, the very function of SATs and PBAs differs tremendously: one is intended to determine mastery within a given school, the other ranks students by overall “aptitude,” most immediately for the sake of college admissions.

It is also difficult to disentangle the effects of certain variables. Since this project is limited to the scope of a single school, conclusions made about the effect of PBA on student performance are heavily influenced by the particular way in which MBA implements PBA. This concern is especially amplified by the fact that PBAs are qualitatively designed and assessed, therefore rendering them more difficult to envision at scale. That said, I still believe this work offers valuable information about the efficacy of PBA and its advantages or disadvantages in comparison with standardized testing. In particular, this analysis interrogates how this form of assessment impacts students’ learning experiences, academic development, college readiness, perceptions of self-efficacy, and ultimately whether PBAs may have a more positive influence on student growth and performance than do standardized tests.

This study considered only one subject matter: social studies. To implement a school-wide performance-based assessment initiative, educators from all disciplines must be prepared to apply these pedagogical designs to their practice. Relatedly, further research should consider performance-based tasks in other content areas (i.e., math, science, art) and the intersections or conflicts of these tasks when a school employs them in concert with one another.
Discussion and Conclusion

This project considered the case study of Metropolitan Business Academy to better understand performance-based assessment. Using a student survey, teacher interviews, and secondary data analysis, I examined how PBA interfaces with instruction and curriculum, and ultimately its efficacy in achieving stated purposes for MBA students. By comparing this form of assessment with the SAT, I framed the trade-offs between standardized assessment and performance-based assessment.

I began with three research questions: Are Metropolitan Business Academy’s performance-based assessments rigorous, or can these data be ascribed to lower standards? Are students benefiting from clearer correspondence among curriculum, instruction, and assessment, or do performance-based assessments and the SAT align similarly? Are students aware of the cognitive processes being probed in performance-based assessments as compared to SAT, or do they understand both assessments to be similar? This study’s findings address each of these questions. Performance-based assessments at MBA are rigorous: educators challenge students and constantly evaluate the assessment standards to ensure validity. Students do benefit from clearer alignment between PBAT and their coursework, and many recalled how these connections reduced test-taking anxiety and unpredictability. Finally, students keenly recognize the different cognitive processes evaluated by each assessment and were more motivated by the inherent rewards of completing a performance-based task.

School culture and Principal Puglisi’s leadership undergird MBA’s system of performance-based assessment. Based on interviews with educators, her vision and directives
were essential to the implementation and eventual success of the program. As one teacher shared, “Judy… knows that changing culture and changing how we teach and how students learn takes time. She never rushes us. She has tremendous energy, but she never makes you feel like you’re behind. She’s just always supportive as we’re trudging along.” This also applied to the overall sense of time demands in the school: “It takes a mindset shift from everybody. If my grades are late, my principal knows we’re in a building where time is elastic. That trickles down to me in how I treat my students. There are parallel processes here in what matters.”

This research is especially important given how it challenges the narrative of underperforming, unmotivated inner-city students. Rather than finding that these students are unmotivated, it seems evident that they are motivated by the highest ends of Bloom’s Taxonomy. When given a performance-based task that affords self-efficacy and opportunities for creation, “underperforming” students seem competitive with peers who earn higher SAT scores. In other words, they perceive the irrelevance or limited learning potential of standardized testing. This may ultimately point to the importance of establishing authentic motivation in any assessment. Greer (2017), for example, examines the connection between motivation and higher attention, which results in maximized learning and performance. These procedures are embedded within performance-based assessments and largely absent from the SAT.

In some sense, this project harkens back to my seventh-grade encounter with the TCAP. We may gain standards of consistency and objectivity by requiring lengthy, multiple-choice

---

18 Performance-based assessment also seemed to benefit educators’ collaborative work. As one teacher shared, “Every year we’re going to revise based on what happened the year before…we are all using the same thing and teaching with similar language/messaging, so we collectively get to see the impact of a well-structured, well-designed lesson module. That improves our self-efficacy since we created this model, we implemented it…and it builds comradery in the design of it, too.”
exams. But what kinds of thinking and learning might we witness in a test that asks students to imagine rather than remember? How might instruction improve when we understand a student’s intellectual journey as a story rather than a norm-referenced data point? In a world increasingly defined by unpredictability, should our praise be awarded to students who perform exactly as we preconceive or who devise their own demonstration of mastery?
Appendix A: Teacher Interview Questions

1. When were you first exposed to the concept of performance-based assessment, and how did you come to learn more?
   a. In-school, via certain administrators, research, colleagues, etc.
2. What were your initial reactions to a concept like PBA, based on your experiences in the classroom?
3. How do you employ PBA in your classroom? What is its role and how do you use it as a tool for teaching or learning?
4. What has been the prevailing attitude of students towards performance-based tasks and assessments? Do any extreme examples (positive or negative) stand out?
5. Do you generally pre- and post-test students throughout the school year? How would you describe the process of assessment in terms of a timeline?
6. How would you compare PBA to the SAT?
7. How do you prepare students for completing a PBA and how does that differ from preparing them for the SAT?
8. Tell me about some of the high-quality projects or portfolios you’ve seen.
9. Tell me about some of the lower quality projects you’ve seen that indicate areas for improvement.
10. Describe the attitudes of students towards their projects at various points along the process (research, editing, presenting, etc.).
11. Have you discussed PBA with parents and community members? How would you characterize those conversations? Is it a relatively approachable topic or a source of confusion?
12. Would you support a widespread implementing of PBA in other high schools in Connecticut? The United States?
13. Can PBA stand in place of standardized testing for legislative or policymaking purposes? Should it?
Appendix B: Student Survey

Survey: Student Views on Assessment at MBA

The multiple-choice questions below ask you to evaluate your experience with different aspects of assessments. Please circle one number between 1 and 5 that best matches your experience.

1. How difficult was the research process involved in your junior year PBAT?
   - Very easy
   - Easy
   - Neutral
   - Difficult
   - Very Difficult

2. How difficult was the presentation involved in your junior year PBAT?
   - Very easy
   - Easy
   - Neutral
   - Difficult
   - Very Difficult

3. How difficult was the peer collaboration involved in junior year PBAT?
   - Very easy
   - Easy
   - Neutral
   - Difficult
   - Very Difficult

4. How difficult was studying for the SAT?
   - Very easy
   - Easy
   - Neutral
   - Difficult
   - Very Difficult

5. How difficult was taking the SAT?
   - Very easy
   - Easy
   - Neutral
   - Difficult
   - Very Difficult

Please circle one number between 1 and 5 that best matches your reaction to the following three statements.

6. My junior year PBAT scores accurately reflect my knowledge and skills.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

7. My SAT score accurately reflects my knowledge and skills.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

8. The SAT and PBAT are basically the same.
   - Strongly disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

Survey: Student Views on Assessment at MBA, cont.

The open-ended questions below ask you to evaluate your experience with different aspects of assessments. Please write a brief answer in response.

9. What did you enjoy about completing your junior year PBAT?
10. What did you dislike about completing your junior year PBAT?
11. What is most different about taking an SAT and taking a PBAT?
12. Would you rather complete an SAT or a PBAT? Why?
13. Do the questions on PBAT seem connected to your classwork and instruction?
14. Do the questions on SAT seem connected to your classwork and instruction?
15. Do you have any other experiences or thoughts to share about PBAT or SAT?

Thank you for completing the survey. Please return this sheet to your teacher and wait quietly until everyone has finished.
## Appendix C: Moderation Study Rubric

<table>
<thead>
<tr>
<th></th>
<th><strong>Exemplary</strong></th>
<th><strong>Competent</strong></th>
<th><strong>Emerging</strong></th>
<th><strong>Novice</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Form an arguable claim/thesis within an introductory paragraph</strong></td>
<td>Take a clear and specific position that can be debated and defended based on thorough knowledge of the context; acknowledge opposing viewpoint(s)</td>
<td>Take a clear and specific position that can be debated and defended based on knowledge of the context</td>
<td>Take a position that can be debated based on a general knowledge of the issue</td>
<td>Give a descriptive position that cannot be debated or defended</td>
</tr>
<tr>
<td><strong>Support and develop claim/thesis using relevant evidence</strong></td>
<td>Select, introduce and explain most relevant quotes and information to support claim/thesis; provide additional context/details for quotes and information</td>
<td>Select and explain relevant quotes and information to support claim/thesis</td>
<td>Select quotes and information to support claim/thesis</td>
<td>Select quotes and information that may not support claim/thesis</td>
</tr>
<tr>
<td><strong>Analysis of evidence</strong></td>
<td>Demonstrate a clear understanding of the evidence; explain how evidence supports thesis; analyze evidence by making inferences, connections and evaluations throughout entire paper</td>
<td>Demonstrate a clear understanding of the evidence; explain how evidence supports thesis; analyze evidence by making inferences, connections and/or evaluations</td>
<td>Demonstrate an understanding of the evidence; explain how evidence supports thesis</td>
<td>Summarizes and/or rephrases evidence</td>
</tr>
</tbody>
</table>
## Appendix D: Moderation Study Planning 2017-2018 (abridged)

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Lesson Content/Materials Needed</th>
<th>Student Deliverable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day One: Tuesday &amp; Wednesday</strong>&lt;br&gt;• Intro to Topic</td>
<td>Guiding Question:&lt;br&gt;• Develop background knowledge about the topic&lt;br&gt;• Political cartoons&lt;br&gt;• Introduce Task and Controversy&lt;br&gt;• Infographic Analysis&lt;br&gt;  o One pager for each infographic (SENT to copy center)</td>
<td>• What do you see? What does it mean? Chart for political cartoon&lt;br&gt;• Analysis of infographics</td>
</tr>
<tr>
<td><strong>Day Two: Thursday &amp; Friday</strong>&lt;br&gt;• Source Analysis&lt;br&gt;• Discussion</td>
<td>• Spectrum activity&lt;br&gt;• Re-teach annotation&lt;br&gt;• Article and analysis</td>
<td>• Source analysis sheet</td>
</tr>
<tr>
<td><strong>Day Three: Monday &amp; Tuesday</strong>&lt;br&gt;• Discussion&lt;br&gt;• Outline (TEAL embedded in outline)</td>
<td>• Discussion&lt;br&gt;  o Handout for discussion protocol with guiding questions</td>
<td></td>
</tr>
<tr>
<td><strong>Day Four: Wednesday &amp; Thursday,</strong>&lt;br&gt;• Finish outline&lt;br&gt;• TEAL Lesson&lt;br&gt;• Review Rubric</td>
<td>• Finish Outline&lt;br&gt;• TEAL Lesson&lt;br&gt;• Review Rubric</td>
<td></td>
</tr>
<tr>
<td><strong>Day Five: Friday, Writing</strong></td>
<td>Writing Day</td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgements

Foremost, I must thank Principal Judith Puglisi and all of the educators at Metropolitan Business Academy who invited me into their school. This paper testifies to their deep concern for students’ learning and well-being.

I am deeply grateful to Professor Richard Hersh for his wisdom and encouragement both during this research project and throughout my time at Yale. Education Studies defined my experience here and I am grateful to Lizzy Carrol, Mira Debs, and William Garfinkel, as well.

Without the community of Education Studies Scholars, this work would be impossible. Thanks to my peers for challenging and teaching me.

Finally, thank you to my family—Mom, Dad, John, Adam, Arrington, and Barb—for investing in my education. I dedicate this work to Pierson and Harrison: may you each create something with your minds that no one can imagine.
Works Consulted


Blömeke, S., & Gustafsson, J.-E. (Eds.). (2017). Standard Setting in Education. New York: Springer International Publishing. [https://doi.org/10.1007/978-3-319-50856-6](https://doi.org/10.1007/978-3-319-50856-6)


