



EXCELLENCE FOR ALL

WORLD-CLASS INSTRUCTIONAL SYSTEMS FOR OUR SCHOOLS

Setting College-Ready Qualification Scores for December 2013 Mississippi World History and Spring 2014 Arizona and Kentucky World History Examinations

The *Excellence for All* initiative calls for students to demonstrate competence in five subject matter areas – mathematics, English language arts, history, the sciences and the arts – to be eligible to earn a proficiency-based diploma (e.g., Arizona’s Grand Canyon Diploma) as early as the end of their sophomore year in high school. Depending on the aligned instructional system their school has chosen, students will take end-of-course examinations in each of these subjects from either the University of Cambridge’s International General Certificate of Secondary Education (IGCSE) or ACT’s *QualityCore* systems. The National Center on Education and the Economy’s (NCEE) Technical Advisory Committee (TAC) has established the qualification scores for English language arts and mathematics for both systems, but the states are responsible for setting the qualification scores in the three other subject areas.

Establishing suitable qualification scores is critical to the success of the initiative because these qualification scores set the standard that every single high school student should be striving for and the standard to which each school is committed to helping every student achieve. At the request of the participating states, NCEE convened a History Task Force to recommend a qualification score for the ACT *QualityCore* and IGCSE History courses. State education agencies in Arizona, Kentucky and Mississippi and the Capitol Region Education Council in Connecticut were each invited to appoint members to the task force, which initially met in May 2012 to set qualification scores for the *QualityCore* U.S. History and IGCSE “World” History examinations. Participants included high school history teachers, community college and university faculty, state education department curriculum professionals, and representatives of the private sector.

In December 2012, NCEE convened a subset of the larger Task Force to recommend a qualification score for the May 2012 administration of the Kentucky World History examination, an assessment that was developed as a complement to the ACT *QualityCore* U.S. History examination used in that state. The World History examination changes yearly and has since been introduced in two other states (Arizona and Mississippi) where schools are also implementing the *QualityCore* instructional system, so NCEE has convened subsets of the larger Task Force on two additional occasions to set qualification scores for the Spring 2013 Arizona and Kentucky examinations. In December 2013, schools in Mississippi administered a world history examination at the end of a semester-long, double-period world history course. This examination will also be used as the end-of-course exam in Spring 2014 for schools in Arizona and Kentucky, where students take world history as a year-long course. In April 2014, NCEE invited previous members of the Arizona and Kentucky world history Task Forces, as well as Mississippi members of

the larger history Task Force, Lee Childress and Richard Damms, to set a qualification score for the December 2013/Spring 2014 administrations of the exam.¹

Because ACT *QualityCore* does not include a world history assessment, teachers at the participating *Excellence for All* schools in Arizona, Kentucky and Mississippi that have adopted *QualityCore* worked together during the course of the 2012-2013 and 2013-2014 school years to assemble an item bank of multiple choice and constructed response questions that could be developed into an end-of-course examination that would be well aligned with the curriculum they were teaching. With this groundwork as a starting point, an NCEE-appointed working committee consisting of Abby Reisman, Columbia University; Emily Wicken and Ann Borthwick, NCEE; and Lloyd Bond, consulting scholar at the Carnegie Foundation for the Advancement of Teaching, worked to create the examination. While the committee determined the format of the exam, the items were based on the historical images and questions selected and designed by the teachers.

The December 2013/Spring 2014 World History examination drew to an extent on the ACT *QualityCore* U.S. History examination to shape its structure, although the examination developers leaned on their own expertise and knowledge of assessment design to construct an examination that would provide reliable and valid information about a student's knowledge and skills in world history. Like the ACT *QualityCore* U.S. History examination, the World History examination includes a multiple choice section and a constructed response section. However, while the constructed response portion of the *QualityCore* examination requires students to answer two constructed response questions (one visual analysis question and one essay question), this World History examination requires students to answer three constructed response questions: two source-based questions (one visual analysis and one text analysis) and one longer essay question.

This examination was designed to include multiple means for students to demonstrate their ability to reason historically, evaluate original source materials, analyze conflicting interpretations of history, and create a historical argument. The intent was for the examination to capture a wider range of skills taught in their courses and would challenge students in ways that have more fidelity with learning history than traditional multiple choice questions typically offer. Because the constructed response section on the World History examination was necessarily lengthier than that of the *QualityCore* examination, the number of multiple choice questions was limited to 20, and the constructed response portion was weighted more heavily (40 points as compared to 20 points in the multiple choice section). The students taking the World History examination have the potential to earn 60 total points: 20 points from the multiple choice section and 40 points from the constructed response section. The constructed response section required students to complete two different types of tasks, including two short answer responses in which students were required to use and evaluate a primary source (each worth 10 points) and a second longer essay in which students were required to make a historical argument (worth 20 points).

¹ A complete set of biographical sketches of the members of the task force is attached as an appendix.

After the administration of the examination in December 2013, the teachers in Mississippi were trained to score the constructed response items using scoring rubrics developed by the same committee that developed the examination. Each item was scored by at least two scorers before a final score was reached. After the results were compiled and analyzed in Winter 2014, the World History Task Force met to review the examination and examples of student work, and to recommend a qualification score for the examination.

The Task Force was charged with two specific goals:

- First, the qualification score should be an indicator of readiness to move forward in education, either to an upper division secondary social science course or to the initial credit-bearing course in a social science program of study in an open enrollment college.
- Second, the score is also intended to serve as an approximation of basic civic competence, indicating that a student who meets the qualification score has demonstrated mastery sufficient to have knowledge consistent with that of an educated person. Students who meet the qualification score should have a set of tools in hand that will allow them to navigate civil society and participate in our democratic institutions.

Because the examination format has changed somewhat from the initial administration in May 2012, there was no extant data on student performance on the examination or on the relationship of student performance on the exam to subsequent success in high school or college. Therefore, the Task Force had to approach developing a qualification score recommendation for this assessment somewhat differently than it had for the two IGCSE History examinations and the *QualityCore* U.S. History examination, though they employed the same approach used to set the qualification score for the previous World History examinations.

The Task Force reviewed examples of student work on the constructed response section of the test, selected to indicate low/middle to high/middle performance on the exam. The Task Force also reviewed a concordance table, prepared by Lloyd Bond, showing the likely performance of students on the multiple choice portion of the exam, given their performance on the constructed response tasks. This process is in accordance with the measurement community's understanding that it is easier for expert panels to come to consensus on student performance levels when they can examine student work rather than relying solely on student answers to multiple choice questions. Dr. Bond reported that student performance on the constructed response tasks appeared to serve as a reasonably good predictor of their results on the multiple choice section. Given this connection, the Task Force decided that a reliable and trustworthy judgment about the appropriate qualification score would be achieved if the Task Force first evaluated student work on the constructed response tasks and then used these findings to help determine the corresponding performance required on the multiple choice items.

Additionally, normative data about the student performance on this examination was tabulated prior to the meeting, but was not distributed to the Task Force until they had come to an initial agreement on the qualification score.

In considering the evidence in hand, the Task Force recognized that the qualification score for World History needed to be seen not just as a mark for a single subject, but as one part of a larger qualification framework in which students would also have to reach a satisfactory score in (i) another history course, (ii) similar standards in the sciences and the arts, and (iii) meet demanding standards in English and mathematics already set by the NCEE TAC. Given the crucial role that English and mathematics play in virtually all post-high school opportunities, the NCEE TAC placed special value on preventing false positives for these subjects (i.e., ensuring that students not ready for college not be told they are). In contrast, the Task Force concluded that for this history exam there was greater value in preventing false negatives (i.e., ensuring that students who could succeed in college were not misclassified by an unreasonably high qualification score and have their path to enrolling in credit bearing community college courses unfairly blocked). Furthermore, while the Task Force shares in the general belief that as a society we need to “raise the bar” for student performance, the Task Force did not believe this exam should be used to create an elite program. There was wide agreement that basic literacy in history was essential to meet the shared objective that students leave high school ready to lead the life of an educated person. Balancing these influences, the Task Force decided that the qualification score for the World History course need not be as stringent as the scores for math and English language arts.

Key considerations for developing a qualification score recommendation for the World History examination were to be found in the answers to these questions:

- What knowledge, skills and dispositions are priorities for this exam?
- How do these priorities align with what is necessary for success in open-enrollment college courses, in upper division high school history courses, and, most importantly, for civic competence?
- What skills reflected in the student work from the constructed response tasks demonstrate readiness for success at open-enrollment colleges or upper division high school history courses?

The Task Force’s deliberations were anchored by five sets of student work drawn from the December 2013 administration. The sets ranged from student performance that the Task Force would characterize as less than adequate to student performance that the Task Force would characterize as representative of the educated person criterion and of readiness to move on to the next level of education.

As a group, the Task Force examined each set of student work, and debated whether the student’s performance (taken as a representation of others performing at a similar level of achievement) demonstrated readiness to move on to the next level of education. Ultimately, they felt that the readiness/educated person benchmark lay somewhere between the fourth and fifth samples of student work, which had earned constructed

response scores of 22.5 and 24, respectively. Although some elements of each sample were slightly higher or lower than necessary, the Task Force concluded that the samples could be taken as a whole and represent satisfactory performance.

After narrowing the discussion to these scores, the Task Force turned to the prediction of multiple choice scores based on the constructed response scores. A score of 22 on the constructed response portion of the test predicts a multiple choice score of 12, as does a score of 23 on the constructed response portion, for overall scores of 34 and 35, respectively. A score of 24 on the constructed response portion of the exam predicts a multiple choice score of 13, for an overall score of 37. Therefore, the Task Force focused in on a range of 34-37 for the qualification score.

There are multiple ways for students to reach the overall score, as points are awarded in a compensatory manner where weaknesses in one set of questions may be offset by strengths in another. A scale score in the range of 34-37 would typically require students to demonstrate roughly a middle range of ability on both the multiple choice and constructed response sections. While it would be possible for a student to earn all of the necessary points on the constructed response section alone, it is unlikely that a student would achieve the qualification score in this way due to the rigor of the constructed response portion of the examination. The Task Force decided that a student who was able to achieve a combined score in this range would have a skill level appropriate to satisfy the Task Force's goals.

Once the Task Force had narrowed their discussion to this range, the normative data for the December 2013 administration of the examination was then shared with them. It indicated that a score of 34 represented the 60th percentile for that administration of the examination, while a score of 37 represented the 70th percentile. The Task Force then considered whether this data might lead them to refine their recommendation in one direction or the other. Cognizant of the fact that the World History examination represents one of eight exams, that students also must meet a qualification score in U.S. History in order to move to the next stage of their education and that false negative decisions should be minimized, but concerned that a score of 34 represented the floor of the range they were considering (because the lower of the two samples they were considering received 22.5 points, it had a hypothetical overall score of 34.5), the Task Force unanimously agreed to recommend a score of **35**.

Over the coming years the *Excellence for All* initiative will gather additional evidence and revisit and make refinements to the qualification score where necessary. Future recommendations will take into consideration the performance of pilot school students in future courses, as well as how they do on college admissions exams as they advance from grade to grade and, eventually, on the grades they receive in college.

April 2014

Appendix: Task Force Members