



# EXCELLENCE FOR ALL

WORLD-CLASS INSTRUCTIONAL SYSTEMS FOR OUR SCHOOLS

## Setting College-Ready Qualification Scores for the Arizona World History Examination

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., a 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 International General Certificate of Secondary Education (IGCSE) or ACT *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 Cambridge 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 2013, NCEE reconvened several of the Arizona members of the larger task force as well as new participants Mark Klobas, instructor of history at Scottsdale Community College, and Jennifer Pesato-Farrlley, Social Studies Specialist at the Arizona Department of Education, to develop a qualification score recommendation for the Arizona World History examination. The Arizona World History examination is an assessment that was designed to serve as a complement to the ACT *QualityCore* U.S. History examination used by some participating schools in Arizona. Because ACT *QualityCore* does not include a world history assessment, teachers at the participating *Excellence for All* schools in Arizona and Kentucky that are implementing *QualityCore* worked during the course of the 2012-2013 school year 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 committee consisting of Abby Reisman, Columbia University; Emily Wicken and Ann Borthwick, NCEE; Brad Fogo, Stanford History Education Group; 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.

After the administration of the examination in May 2013, the teachers 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 Fall 2013, the Arizona 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 passes has demonstrated mastery sufficient to have knowledge consistent with that of a reasonably 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.

The qualification score is *not* intended to approximate readiness for a career as a historian or other social scientist, nor for entrance into a social science program at a selective postsecondary institution. The task force assumed that such college- or career-readiness paths would demand a higher benchmark than “literacy.”

The students taking the Arizona 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 one set of short answer responses in which students were required to use and evaluate two primary sources and a second longer essay in which students were required to make a historical argument. Each task is worth up to 20 points. Because the examination was brand new, there was no data available to correlate student performance on the examination with subsequent success in high school or college. Therefore, the Task Force had to approach its qualification score recommendation for this assessment somewhat differently than the larger History Task Force had for the IGCSE “World” History and *QualityCore* U.S. 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 was because the measurement community has come to understand 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 more 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 the examination was tabulated prior to the meeting, but was not distributed to the Task Force until the Task Force had come to an initial agreement on an appropriate qualification score range.

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) in the sciences and the arts, and (iii) in English and mathematics, which have already been set by the NCEE TAC. Given the crucial role that English and math 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 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 Arizona 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 Arizona World History examination were 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 May 2013 constructed response tasks demonstrate readiness for success at open-enrollment colleges or upper division high school history courses?

The Arizona World History examination drew to an extent on the ACT *QualityCore* U.S. History examination to shape its structure, although the examination developers followed their own expertise and knowledge of assessment design to construct an examination that

is expected to provide reliable and valid information about a student's knowledge and skills in world history. Like the ACT *QualityCore* U.S. History examination, the Arizona 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), the Arizona 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.

Introducing this greater degree of variability into the assessment along with 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 was designed to yield an examination that would represent the 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 Arizona 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 Task Force's deliberations were anchored by examples of student work drawn from the May 2013 administration. Five completed constructed response sections were analyzed by the Task Force to determine the qualification score. The examples ranged from student performance that the Task Force would characterize as less than adequate to student performance that the Task Force would characterize as quite satisfactory.

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, the Task Force concluded that the readiness benchmark lay somewhere between the third, fourth and fifth examples of student work, which had earned constructed response scores of 20, 22 and 24 points respectively. Although some elements of each of the samples were slightly higher or lower than necessary, the Task Force concluded that the samples could be taken as a whole and represent reasonable 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 20 on the constructed response portion of the test predicts a multiple choice score of 14, for a total score of 34. A constructed response score of 22 predicts a multiple choice score of 15, for a total score of 37, while a constructed response score of 24 also predicts a multiple choice score of 15, for a total score of 39. Therefore, the task force focused on a range of 34-39 for the qualification score.

Students may reach the required scores in multiple ways, since points are awarded in a compensatory manner where weaknesses in one set of questions may be offset by

strengths in another. The Task Force concluded that a scale score in the range of 34-39 would typically require students to demonstrate an adequate range of ability on both the multiple choice and constructed response sections. While it would be theoretically 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.

Taking all of these factors into account, the Task Force began with a tentative qualification score recommendation of 38. NCEE staff then shared the normative data for the May 2013 administration of the examination with the Task Force. Based on the May 2013 test results, a score of 38 represented the 99<sup>th</sup> percentile for that administration of the examination. The Task Force considered this data and decided to refine its recommendation. Cognizant of the fact that the World History examination represents one of eight exams and that students also must meet a qualification score in U.S. History in order to become eligible for a Grand Canyon Diploma, the Task Force concluded that a 38, at the 99<sup>th</sup> percentile, would likely result in false negative decisions (that is, a qualification score this high would likely exclude students who were in fact ready to succeed in community college or in higher level high school social science courses), and decided to adjust the qualification score recommendation to the lower end of the earlier agreed upon range. The Task Force unanimously agreed that **34** should stand as its recommendation, noting that even at this level the required performance was in the 94<sup>th</sup> percentile.

Given that the 2013 administration was the first year this examination was administered in Arizona, that the sample size of students was relatively small, and that the exact nature of the examination was not known to either teachers or students until well into the school year, the Task Force wished to emphasize that this first qualification score recommendation should be viewed as provisional. When more schools are participating in the initiative and schools have more experience with the curricula, students are likely to become more familiar with the subjects covered in the examination. As a result, the quality of data is likely to improve significantly and allow a more reliable benchmark to be established. Consequently, beginning next year and over the subsequent years the *Excellence for All* initiative, at the state's request, will gather additional evidence and suggest refinements to the qualification score where appropriate. Future recommendations will take into consideration the performance of pilot school students in future courses, as well as how students who take the examination perform on college admissions exams as they advance from grade to grade and, eventually, what grades they receive in college.

*January 2014*

Appendix: Task Force Members

## **Arizona World History Task Force Members**

### **Mark Klobas**

Mark Klobas is an instructor of history at Scottsdale Community College in Scottsdale, Arizona. A graduate of the University of Arizona, he received his Ph.D. in history from Texas A&M University. He has developed curriculum for a range of courses, from American history and world history, to specialized courses on English history, Irish History, and the history of the modern Middle East.

### **Jennifer Pesato-Farrlley**

Jennifer Pesato-Farrlley serves as the Arizona Department of Education Social Studies Specialist. She has twelve years of experience teaching World History, AP World History, and AP European History at the high school level. She has previously served as a district-level K-12 Director of Curriculum, and is in the process of completing an Ed.D. in curriculum and instruction from Northern Arizona University.

### **Kevin Olson**

Kevin Olson is a partner in the law firm of Steptoe & Johnson LLP, a Washington D.C. firm with offices in Phoenix, Los Angeles, New York, Chicago, London, Brussels, and Beijing. He practices in the firm's business solutions group. His work is in the areas of general corporate advice, start-up and later stage financings, securities and corporate finance, mergers and acquisitions, and other commercial transactions. Mr. Olson represents clients in many industries, including consumer electronics, alternative energy, food distribution, and technology. He is also a member of the Securities Section of the State Bar of Arizona (for which he is a past chairman) and the Business Section of the State Bar of Arizona (for which he is also a past chairman).

### **Christine Wilfong**

Christine Wilfong teaches social studies at Dysart High School in Phoenix, where she also chairs the social studies department. She has also chaired the district social studies curriculum committee and the most recent textbook adoption process. Ms. Wilfong led the committee that produced the first district benchmarks for social studies and the first shared 9-12, district-wide common curriculum for grades 9-12, and pioneered the DBQ process for grades 6-12. She also helped to produce the district's first online curriculum and website for district teacher resources. Ms. Wilfong holds a B.A. in History, an M.A. in Curriculum and Instruction and Administration and Supervision, and highly qualified certificates in Geography, History, Government and Economics, along with Principal and Supervision.

## **Advisors**

### **Lloyd Bond**

Lloyd Bond is a consulting scholar with the Carnegie Foundation for the Advancement of Teaching and emeritus professor of education at the University of North Carolina, Greensboro. From 2002 to 2008 he was a senior scholar at Carnegie working in the area of assessment across several Carnegie Foundation programs. Dr. Bond has published widely in the area of assessment, measurement theory and testing policy and has made fundamental contributions to the literature on measuring complex performance and cognitive process underlying test performance. He has held editorial positions on the leading journals in educational and psychological measurement and serves on numerous commissions and panels devoted to testing and testing policy. He is currently a member of the Data Analysis Committee of the National Assessment of Educational Progress (NAEP) and the Psychometric Panel of The College Board. Previously he served on the National Academy of Sciences' Committee on Indicators of Science and Mathematics Education and their Committee on Science Assessment Standards. A fellow of both The American Psychological Association (APA) and the American Educational Research Association (AERA), he is the recipient of numerous honors and awards, including the Presidential Citation from AERA for Contributions to Educational Measurement and an APA Distinguished Service Award for his work on the Joint Standards for Educational and Psychological Testing. He has served as a trustee for The College Board, and currently sits on the boards of the Human Resources Research Organization and the National Center for Research on Evaluation, Standards & Student Testing. Dr. Bond obtained a Ph.D. in Psychology from the Johns Hopkins University, specializing in psychometrics and quantitative methods. He taught test theory and psychometrics at the University of Pittsburgh, and at the University of North Carolina, Greensboro.

### **Avishag (Abby) Reisman**

Avishag (Abby) Reisman is a visiting professor at Teachers College, Columbia University. She was formerly a senior researcher at the National Center for Research on Evaluation, Standards, & Student Testing (CRESST) at UCLA. Dr. Reisman was the project director for "Reading Like a Historian" in San Francisco, the first extended history curriculum intervention in urban high schools. Her work on the project has appeared in *Cognition and Instruction* (2012) and the *Journal of Curriculum Studies* (2012). Her dissertation, which emerged from this project, won the 2011 Larry Metcalf Award from the National Council of the Social Studies. She played a key role in the research and development of "Historical Thinking Matters," and designed the history curriculum website for the Bill Lane Center for the Study of the North American West.