Appendix B. Template for Quality Verification of Testing Instruments

To establish the adequacy of the testing instruments used in response to the RFA, download and complete the Template for Quality Verification of Testing Instruments provided on the CDE SES RFA Web page at http://www.cde.ca.gov/fg/fo/r16/ses13rfa.asp.
The template is included in the 50 page application maximum.

Name of Applicant Entity:
Accel Online

Name of Test to be Verified

California Standards Test (CST)

Publisher of Test

California Department of Education

Applicants must appropriately address the three topics in the template:

Test Rationale

State the purpose of the test
Indicate the content and skills to be tested
Indicate the intended test takers (i.e., grade-level and subject)

Appendix B. **Purpose**. Response:
The purpose of the test to determine the student's level of understanding of the standardized content deemed appropriate by the State of California Department of Education for that student's grade level in mathematics, language arts or science. The CSTs, given in English, are designed to show how well students in grades two through eleven are performing with respect to California's content standards. These standards describe what students should know and be able to do at each grade level in selected content areas.

Appendix B. **Content and Skills to be Tested**. Response:
The content tested is a set of standards for each grade level in mathematics, language arts and science that is set by the department of education. This content is the same for every child in the state at any one grade level and in each subject. The skills are those the department of education expects children of each grade level to have mastered before being allowed to promote to the next grade. CSTs are administered in four content areas: ELA, mathematics, history–social science, and science.

Appendix B. **Test Takers**. Response:
The test takers are students in grades 3-12 in the state of California who are taking a math, language arts or science class this year.
Note: Type responses in this table. As the response is entered, this box will grow.

Technical Qualities of the Test

Provide evidence describing the technical qualities of the test. Evidence must include reliability and validity analyses or studies that determine whether the test meets its intended purpose.

At a minimum, content **validity** must be established.

If there was a review of the test by a panel of experts on curriculum in the subject that the test is intended to measure, describe the procedures for review and the qualifications of the panel.

If there was a study or analysis conducted to determine validity, describe what type of study or analysis was conducted and the results of the study or analysis.

Appendix B. **Procedures Used and Panel Qualification**. Response:
Validity refers to the degree to which each interpretation or use of a test score is supported by evidence that is gathered (AERA, APA, & NCME, 1999; ETS, 2002). It is a central concern underlying the development, administration, and scoring of a test and the uses and interpretations of test scores.
Validation is the process of accumulating evidence to support each proposed score interpretation or use. It involves more than a single study or gathering one particular kind of evidence. Validation involves multiple investigations and various kinds of evidence (AERA, APA, & NCME, 1999; Cronbach, 1971; ETS, 2002; Kane, 2006). The process begins with test design and continues through the entire assessment process including item development and field testing, analyses of item and test data, test scaling, scoring, and score reporting.

Appendix B. **Type of Study or Analysis and Results**. Response:
According to The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), analyses that demonstrate a strong relationship between a test's content and the construct that the test was designed to measure can provide important evidence of validity. In current K–12 testing, the construct of interest usually is operationally defined by state content standards and the test blueprints that specify the content, format, and scoring of items that are admissible measures of the knowledge and skills described in the content standards. Evidence that the items meet these specifications and represent the domain of knowledge and skills referenced by the standards supports the inference that students' scores on these items can appropriately be regarded as measures of the intended construct.

Correlations Between Scores on the CSTs and Scores on the CAT/6 Survey Convergent validity evidence was collected in 2004 by examining the relationship between CSTs and their CAT/6 Survey (Terra Nova, 2nd Edition, 2000) counterparts. The CAT/6 Survey is a norm-referenced test that assesses students in reading, language, spelling, mathematics, and science and evaluates student achievement in terms of norms. The CSTs were expected to relate closely to their counterparts in the CAT/6 Survey programs when they measured similar constructs, and to correlate less well when they measured different constructs. A full description of the study can be found in the California Standardized Testing Program Technical Report, Spring 2005 Administration, linked on the CDE STAR Technical Reports Web page at http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp. A summary of findings follows:
Correlations Between Scores on the CST for ELA and Scores on the CAT/6 Survey

Reading/Language/Spelling—The study showed that, as expected, CST for ELA scores in all grades correlated highly with scores on both the CAT/6 Survey Reading Language tests, because these tests assessed similar skills. The correlation coefficients between the CST for ELA and CAT/6 Survey Spelling tests were somewhat lower, which is to be expected because these tests measured somewhat different skills.

Correlations Between Scores on the CST for Mathematics and Scores on the CAT/6 Survey Mathematics —In grades two through seven, student scores on the CST Mathematics tests correlated highly with their scores on CAT/6 Survey Mathematics test.

Results of the validity correlation tests follow.

Note: Type responses in this table. As the response is entered, this box will grow.

At a minimum, test **reliability** must be established.
If there was a study or analysis conducted to determine reliability, describe the type of study or analysis conducted (test-retest, internal consistency, etc.).
Describe the results of the study or analysis.

Appendix B. **Study or Analysis Conducted**. Response:
Reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested, rather than fluctuations due to chance or random factors. The variance in the distribution of test scores—essentially, the differences among individuals—is partly due to real differences in the knowledge, skill, or ability being tested (true-score variance) and partly due to random unsystematic errors in the measurement process (error variance).
The number used to describe reliability is an estimate of the proportion of the total variance that is true-score variance. Several different ways of estimating this proportion exist. The estimates of reliability reported here are internal-consistency measures, which are derived from analysis of the consistency of the performance of individuals on items within a test (internal-consistency reliability). Reliability coefficients may range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores if they were retested.

The formula for the internal consistency reliability as measured by Cronbach's Alpha (Cronbach, 1951).
Appendix B. **Results of Study or Analysis**. Response:
A study was performed on all 38 individual CST exams with the average alpha score of 0.94 showing that the CST is a highly reliable exam. Overall reliability—The reliability analyses on each of the 38 operational CSTs are presented in Table 8.2. The results indicate that the reliabilities for all grade-level CSTs for ELA, history–social science, mathematics, and science were very high, ranging from 0.92 to 0.95. Reliability estimates for the EOC CSTs also tended to be high with the exception of the CST for Integrated Mathematics 1 and 2 and the Integrated/Coordinated Science tests, which were in the mid- to high-0.80s.

Note: Type responses in this table. As the response is entered, this box will grow.