



Action Learning Systems, Inc.

English-Language Arts Benchmark Tests

2004 Technical Report

Action Learning Systems, Inc.

174 W. Foothill Blvd #306

Monrovia, Ca 91016

Telephone/ (626) 357-8041

Fax/ (626) 357-5031

TABLE OF CONTENTS

| | |
|--|----|
| INTRODUCTION | 2 |
| DEVELOPMENTAL FRAMEWORK OF THE BENCHMARK TESTS | 2 |
| Description of the Content Standards Measured | 3 |
| Development of the Benchmark Tests | 4 |
| Validity of the Benchmarks Tests | 4 |
| Description of the Performance Levels | 5 |
| Administration of the Benchmark Tests | 5 |
| ENGLISH-LANGUAGE ARTS BENCHMARK TESTS | 5 |
| Demographics of Participants in Form 3 Test Administration | 6 |
| Relationships between Benchmark Tests and CST | 7 |
| Benchmark Tests and Annual Measurable Objectives | 8 |
| CONCLUSION | 8 |
| Future Steps | 9 |
| Acknowledgments | 9 |
| APPENDIX – Benchmark Reports | 10 |

INTRODUCTION

On Jan. 8, 2002, President Bush signed into law the *No Child Left Behind Act of 2002 (NCLB)*. This new law represents the federal government's most extensive restructure of the 1965 Elementary and Secondary Education Act (ESEA). This act incorporates the principles and strategies proposed by President Bush. These include increased accountability for states, school districts, and schools; greater choice for parents and students, particularly those attending low-performing schools; more flexibility for states and local educational agencies (LEAs) in the use of Federal education dollars; and a stronger emphasis on reading, especially for our youngest children. The state of California's NCLB accountability plan embraces high quality tests aligned with state adopted standards. Action Learning Systems Inc. (ALS) addresses NCLB legislation and state standards with its development of curriculum-aligned, formative benchmark tests.

Through its creation of content, and performance standards for English-Language Arts and Mathematics, the state of California has defined what a student should know and at what level of proficiency. Through the adoption of these standards, the state is clearly affirming what content students need to acquire at each grade level. With these standards in place, student achievement and mastery of these standards are measured with the California Standards Tests (CST), criterion-referenced tests developed specifically for California. As part of the state's accountability system, performance on the CST also constitutes the largest component of a school's API (Academic Performance Index).

Research has consistently shown that the use of formative tests (i.e., benchmark tests) is a strongly recommended method to gauge mastery throughout the school year, provide teachers with diagnostic and prescriptive information, and provide students with test-taking skills. To assist districts, schools, and teachers, ALS has implemented a focus on these standards through formative benchmark testing.

DEVELOPMENTAL FRAMEWORK OF THE BENCHMARK TESTS

ALS has developed Benchmark tests to measure student progress in mastering the California Standards in specific grades and/or subjects. The term benchmark was adopted

to emphasize the concept of on-going assessment throughout the year at key instructional points prior to the annual administration of the state’s on-demand assessments.

Description of the Content Standards Measured

The State Board adopted the California English-Language Arts content standards in November of 1997 and the mathematics standards in December of 1998. These standards designate the content to be taught and what all students should be proficient in by the end of each grade level in the respective content areas. The NCLB legislation requires that all students be at or above “proficient” in these two content areas by the year 2014. In California, “proficient” or above is determined by performance on the CST. In addition, prior to 2014, schools are to have met designated annual measurable objectives (AMO), i.e., percentage of students required to be at “proficient” or above.

At each grade level, there are numerous standards designated to be taught in one school year. With these large numbers, the requirement that all students master all standards and reach the levels of “proficient” or above on these high stakes tests is unrealistic. For example, in English-Language Arts the number of standards to be taught in grade four is 67; at grade seven it is 65; and at grade nine is 107. In response to the impracticality of teaching and mastering these large numbers of prescribed standards, ALS has selected a smaller number of standards that are considered essential for each grade level and/or mathematics discipline. These essential standards, also referred to as “power” or “focus standards,” were selected by content area experts after a thorough review of state standards, the determination and weighting of the most tested items on state tests, and instructional sequence.

Focus Standards

ALS engaged a representative group of in-service teachers and curriculum specialists to identify by grade level and subject, the most salient and/or important standards for devoting instructional emphasis. These were identified as “focus standards.” Once the focus standards were determined, blueprints were developed for each grade and/or course. All blueprints are directly aligned with the CST and are reviewed by teams of content-area experts.

Development of the Benchmark Tests

Teams of professional item writers developed the format and items for the ALS Benchmark tests as prescribed by the CST blueprints. The teams were designated by content area and each member had several years of experience in developing test items for standards-based state assessment programs (e.g. Golden State Exam, CLAS, CST, and CAHSEE). Lead team members also held lead or supervisory positions in state test development programs.

Team members reviewed grade appropriate textbooks for item development. Prior to developing the math items, the teams reached agreement on the instructional pacing sequence by content by grade. Each focus standard was assessed with a minimum of three items. Each item had four distracters. English-language arts reading passages and mathematics word problems were selected by grade level and length that corresponded to those utilized on the California standards-based assessments.

All Benchmark tests undergo a field-testing process. Item analyses are performed, which include meticulous analyses of p-values, pt. biserial coefficients, and other indices of discrimination, after each round of field-testing. The acceptability of difficulty levels included that percent correct in the 30 percent to 80 percent range. The discrimination level for each item was at or above 0.3. Items not meeting psychometric criteria are either eliminated and replaced, or modified. New versions are subsequently field tested until the complete test has been determined psychometrically sound.

Validity of the Benchmark Tests

English-Language Arts and mathematics subject experts were selected to participate in a validation study. These experts carefully reviewed all items on each benchmark test to determine how well they measured the “focus standards.” Each item was reviewed for alignment to content-grade-specific focus standard, instructional validity (i.e., appropriate grade-level vocabulary and sentence structure, etc.), and bias (i.e., gender, ethnic, offensive language or situations, etc.). This validation review met with the standards as outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999).

Description of the Performance Levels

ALS Benchmark tests and the CST are both criterion-referenced assessments. As such, these tests compare a student's score with a common standard of performance. Percent-correct scores determine whether a student has established minimum acceptable performance. The ALS Benchmark test results are reported using the same performance levels as are used with the CST (i.e., Advanced, Proficient, Basic, Below Basic, and Far Below Basic).

Future analyses will determine the statistical relationships between the performance level designations on the Benchmark tests to the CST designated performance levels. At this time, the Benchmark performance levels should be used as a formative indicator and not an exact predictor of a student's CST performance.

Administration of the Benchmark Tests

As a formative tool, the series of Benchmark tests are sequenced, for the most part, to be administered over the course of the school year, prior to the annual spring administration of the CST. It is intended that the tests be administered in one normally scheduled class period. Students taking those classes for which Benchmark tests have been developed should take the tests in that class. All benchmark tests assess standards with multiple-choice questions. Each question contains one item stem and four distracters. Students record their response to each question on a separate answer sheet (i.e., scantron sheet). Overall, Benchmark tests should not take more than 40 minutes each to administer. To maintain validity and reliability of the assessment results, it is critical that the Benchmark tests remain intact – meaning that the items written and assessed in a specific order or format.

ENGLISH-LANGUAGE ARTS BENCHMARK TESTS

Three forms of both the English-Language Arts Benchmark tests were developed by 2004. The English-Language Arts Benchmark Tests - Form 1 and Form 2 were administered during the 2002-2003 school year. Subsequent analyses led to further refinement of the items included in the administration of the Form 3 series during the 2003-2004 school year.

Demographics of Participants in Form 3 Test Administration

Students from the Salinas Union High School District (SUHSD) and the Stockton Unified School District participated in this test administration. Demographics of the student populations of these two districts are presented in Table 1.

Table 1.

| Demographics based on August 31, 2004 AYP data release | | | | |
|---|--------------|-----------------|--------------|----------------|
| Characteristic | SUHSD | Stockton | Total | Percent |
| ALL STUDENTS | 6762 | 22671 | 29433 | --- |
| Ethnic Subgroups: | | | | |
| African-American | 159 | 2952 | 3111 | 11 |
| Indian | 21 | 913 | 934 | 3 |
| Asian | 115 | 3019 | 3134 | 11 |
| Filipino | 247 | 1172 | 1419 | 5 |
| Hispanic | 5324 | 11810 | 17134 | 58 |
| Pacific Islander | 14 | 118 | 132 | 0 |
| White | 871 | 2655 | 3526 | 12 |
| Socio Economic Disadvantaged | 4211 | 16960 | 21171 | 72 |
| English Language Learners | 3805 | 8490 | 12295 | 42 |
| Students with Disabilities | 579 | 1968 | 2547 | 9 |

By ethnic subgroups, the student population consisted of approximately fifty-eight percent Hispanic; twelve percent white; eleven percent African-American; eleven percent Asian; and 5 percent Filipino. Seventy-two percent of the students were classified as “Socio Economic Disadvantaged” (as indicated by participation in the free/reduced lunch program), and forty-two percent were classified as English Language learners.

From this larger population, a total of 15,898 English-Language Arts Benchmark tests - Form 3 were completed by students in grades 7 through 11 during the 2003-2004 school year. Table 2 presents the number of tests administered by grade level.

Table 2.
Total Number of Tests Administered by Grade Level

| Grade | # |
|-------|-------|
| 7 | 3389 |
| 8 | 4426 |
| 9 | 3055 |
| 10 | 2686 |
| 11 | 2342 |
| Total | 15898 |

Relationships between Benchmark Tests and CST

A major premise in the development and use of ALS benchmarks is that there is a positive relationship with scores a student receives on both the ALS benchmark tests and the California Standards Tests. One way of expressing this relationship, for example is, if a student scores high on the ALS Benchmark test, that student should also score high on the CST. The data shows, in fact, there are strong positive correlations (ranging from .65 to .81) between the ALS English-Language Arts Benchmark Tests - Form 3 (the series of four sequenced components) and the 2004 English-Language Arts CST at each grade level measured. Table 3 displays the breakdown of correlations for each sequenced component of Form 3 by grade level. The numbers in parentheses represent the numbers of students that had a completed test and a CST score. BM1 through BM4 represent the four components that make up the Form 3 series.

Table 3.
Correlations between ALS ELA Benchmark Tests – Form 3 and 2004 ELA CST

| Grade | BM1 | BM2 | BM3 | BM4 |
|-------|---------------|---------------|---------------|---------------|
| 7 | .746 (n=1984) | .776 (n=1474) | .739 (n=1576) | .730 (n=2064) |
| 8 | .696(n=2678) | .710 (n=1946) | .694 (n=2141) | .723 (n=2435) |
| 9 | .645 (n=894) | .787 (n=1505) | .808 (n=614) | .763 (n=1163) |
| 10 | .754 (n=825) | .749 (n=1242) | .763 (n=1550) | .740 (n=1117) |
| 11 | .766 (n=675) | .769 (n=1008) | .770 (n=613) | .750 (n=858) |

As may be seen in Table 3, the correlations between the Benchmark tests and the CST for this group of students was strongly correlated. These results lend support to the capability of the Benchmark tests in predicting CST performance. Future research and analyses will focus in this area.

Benchmark Tests and Annual Measurable Objectives

A major component of the federal government's No Child Left Behind (NCLB) legislation signed by President Bush in 2002 is that all students will attain "proficiency" in reading and mathematics by 2014, including students with disabilities and English learners.

California's provision of accountability under NCLB addresses the foregoing component by establishing adequate yearly progress (AYP) goals as determined by annual measurable objectives (AMO), participation rate, academic performance index (API), and high school graduation rate. The annual measurable objectives (AMO) at the elementary and middle schools are based on the CST in English-Language Arts and mathematics; and the California Alternate Performance Test (CAPA). At the high school level, the AMO's are based on results on the California High School Exam (CAHSEE) and the CAPA.

As prescribed by the state performance levels, mastery is considered at "proficient" or above. NCLB requires that a specific percentage of all students meet this level of proficiency each year. Upon each administration of the Benchmark test, results show those objectives either being mastered or not being mastered by the student. The strong positive relationship between ALS' ELA Benchmark Tests – Form 3 and the ELA CST test, one of three tests used to calculate AMO, contributed considerably to schools attaining their targets.

CONCLUSION

ALS' focus on standards and benchmark tests aligned with the standards has emerged as a result of state and federal requirements (i.e., NCLB). It is expected that through the use of benchmark testing students will come to expect and demand meaningful assignments with clear purposes, i.e., standards-based. They will understand the idea of looking at exemplars to help them understand the quality of work expected of them. Teachers will develop units that must be organized around standards. Teachers' activities

will be justified in terms of standards. Teachers will use benchmark test results as formative tools as they prepare students to learn how to reason, apply knowledge, and produce quality work. ALS Benchmark Tests carefully aligned to clear instructional objectives can be a means of raising student motivation and achievement. The student test cycle is critical if students are to perform at higher levels.

Future Steps

In the effort to provide the most thorough information regarding test development, reliability, and validity, steps are currently being taken to provide technical information for all Benchmark tests currently developed by Action Learning Systems, Inc.

Acknowledgements

Action Learning Systems, Inc. would like to recognize the contributions of Linda Murai, former Director of Performance Assessment of Sacramento County Office of Education, and Dr. Robert Martinez of Salinas Union High School District for their guidance, psychometric expertise, and analytic work on this report. ALS would also like to acknowledge the wonderful contributions and expertise of all members of the test development team for their past, current, and future efforts.

APPENDIX

Benchmark Test Reports

In a joint venture with *Achieve Data Solutions LLC—Data Director*, three distinct Benchmark reports have been developed and can be available for use by students, teachers, and administrators:

- The **Student Exam Report** includes the response made for each question, and related standard, noting whether the response was correct; the number and percent correct; the student's performance level; and the number correct for each standard.
- The **Classroom Exam Report** developed for each teacher's classroom, includes the frequency of response for each multiple-choice item and standard; the correct response for each question; the average number and percent correct for the classroom; the number of students in each performance level; and the number and percent of students answering each specific standard correctly.
- The **School Exam Report** includes, school wide, the percent correct for each classroom's result by standard; the overall percent correct for each standard; and the number of students in each performance level.

Each report provides information on the performance level attained either by student, classroom or school. With the classroom and school reports, the annual measurable objective (AMO) rate may be calculated. Samples of all three of these *Data Director* reports begin on the following page.

