

California Association of Independent Schools
2007 ERB Workshop Schedule
Presented by Dr. Sid Barish

9:00-10:30 a.m.

Introduction

Overview of CTP 4
Value of Formative Assessment
How to read CTP 4 Reports
Interpretation & Use of CTP 4 Score Reports

10:30-12:00

Analysis of Case Study Reports (practice what you learned)

- review of CTP 4 score reports
- examining scores
- determining relevance of scores
- tips on explaining results to stakeholders

12:00-1:00

Lunch

1:00-2:30

Use of *TestWiz* to Track Student Achievement

- introduction of Web version of TestWiz.net
- selecting and viewing reports
- customizing reports & parameters

2:30

Closing remarks



Educational Records Bureau

Dear Colleague:

ERB frequently receives questions about the content, interpretation, and use of the CTP 4. In an effort to help its member schools understand the purpose and value of the test battery, ERB offers the following information. For more information in any of these areas, please go to the *CTP 4 Content Standards Manual*, *CTP 4 Handbook for Classroom Teachers*, *CTP 4 Handbook for Administrators*, and CTP 4 Technical Report available on the CTP 4 order form.

-
1. *What kind of test is the CTP 4?*
 2. *How were the standards determined for "meeting," "exceeding," or "developing" in the Level 1 and 2 tests?*
 3. *Can we test over an extended period of time?*
 4. *Is a shorter version of CTP 4 possible?*
 5. *How can I learn to interpret CTP 4 results to individualize instruction more effectively?*
 6. *What are the best ways to use the test results?*
 7. *How should we interpret discrepancies between achievement and reasoning results?*
 8. *How can we use the item analysis results as a helpful tool?*
 9. *How can we help parents interpret test results and understand norm group comparisons?*
 10. *What are percentiles and stanines?*
 11. *Are workshops available for teachers on ERB tests?*
 12. *Is there a help line available from ERB?*
 13. *Is the CTP 4 useful as an admission test?*
 14. *What help can ERB give in easing test anxiety?*
 15. *What is the best way to view CTP 4 trends?*

Q: What kind of test is the CTP 4?

A: The CTP 4 is a test battery consisting of a series of multiple-choice and open-ended questions administered to groups of students over the course of several days. CTP 4 has 10 levels, with a different level used with each grade from 1-11. Students in the early grades (levels one and two) take the test under untimed conditions and mark their answers in the test booklet. Students in the middle and upper grades take the CTP 4 under timed conditions and mark their answers on a separate answer sheet. This is not a power test and most students will be able to complete the test in the time allowed.

All 10 levels of the CTP 4 include achievement tests that measure what the student has learned in the areas of reading and mathematics. In addition, levels 1 and 2 test word

analysis and listening skills. Beginning with level 3 there are also tests that measure verbal reasoning, quantitative reasoning, certain writing skills, and vocabulary.

Q: How were the standards determined for “meeting,” “exceeding,” or “developing” in the Level 1 and 2 tests?

A: The number of questions in a content area varies. In most content areas, there are 8 to 13 questions. The number of correct answers required to meet expectations and the number required to exceed expectations, in each content area of each test, are based on the judgments of a panel of teachers. There were two panels, each consisting of seven teachers of first-graders and second-graders at schools using CTP tests. One panel made the judgments for the reading, listening, and writing tests; the other panel made the judgments for the mathematics tests. Before making the judgments for each content area of each test, the panelists examined the questions carefully. They looked at statistics describing the performance, in that content area, of the students who took the test at the special equating administration; they also looked at statistics showing how that group of students compared with the suburban and independent norm groups. Each panelist, working individually, then made an initial pair of judgments, specifying the number of correct answers that were required to meet expectations and the number required to exceed expectations. The session leader (not one of the panelists) tabulated these initial judgments and displayed the results, and the panelists discussed their judgments and the reasoning behind them. Finally, each panelist, again working individually, made a second pair of judgments. In most cases, these judgments revealed a clear consensus. In a few cases, the panelists were divided almost evenly between two adjacent score levels, and one of those two score levels was selected, based on the performance of students responding to those items in the equating administration. The panel included educators from the following schools:

Radnor Township School District (Bryn Mawr, PA)
Providence Day School (Charlotte, NC)
Merion Elementary (Merion, PA)
Quest Academy (Palatine, IL)
Kingswood Academy Hinsdale, IL)
Stuart Hall for Boys (San Francisco, CA)
Fairfax Elementary (Fairfax, VA)
Maimonides Academy Los Angeles, CA)
Foothill Country Day School (Claremont, CA)
Tower School (Marblehead, MA)
Heathwood Hall (Columbia, SC)
T.W. Miller School (Norwalk, CT)

Q: Can we test over an extended period of time?

A: Yes. Schools may select their own test schedule. The only restriction ERB has is that once students begin a test they must complete that section. Otherwise, schools may schedule tests in any order and for amounts of time they think best for their students. ERB finds that most schools test parts of 3-5 days for 1-2 hours per session, depending on the age of the children.

Q: Is a shorter version of CTP 4 possible?

A: It is not necessary to administer the entire test battery. Schools may elect to use as many parts of the CTP 4 battery as they wish, and scores will be provided only for those content areas taken and sent in for scoring.

Q: How can I learn to interpret CTP 4 results to individualize instruction more effectively?

A: The *CTP 4 Content Standards Manual* contains helpful information about the scope and sequence of the tests along with content category information and sample questions. Similarly, the *CTP 4 Handbook for Classroom Teachers* and *CTP 4 Handbook for Administrators* offer sample reports and a guide to interpreting them. In addition, the handbooks offer suggested lessons and activities to teach the content and skills covered by the tests at various levels. It may be advantageous to contact an ERB test consultant for a workshop and then apply the information and skills to your school and students.

Q: What are the best ways to use the test results?

A: It is best to compare the same cohort of students from year to year (as opposed to, say, grade 7 one year and another group of 7th graders the next year). Scale scores are a good indicator of growth from year to year within the same content area. Typically, scale scores will increase 6-8 points in verbal areas and 8-12 points in math areas, from year to year, for the same group of students.

The Administrator's Summary will help you compare your students' achievement to national, suburban public, and independent schools. The report helps you determine the percentage of students at your school (the "local" score) above average, average, and below average against what is "expected" in each category and in comparison to the three norm group populations. By using that information alongside other work the students are doing daily, the schools may identify areas of strength and weakness.

The Item Analysis provides achievement information for individual and groups of students to inform instruction by content and areas within that content. Patterns of weakness in an area of the curriculum as measured by the CTP 4 may have implications for work on curriculum and point to needed intervention for students.

Q: How should we interpret discrepancies between achievement and reasoning results?

A: ERB generally recommends that schools view 2-stanine differences between reasoning and achievement tests in related areas as indicators of differences that should be investigated. High reasoning scores accompanied by low achievement scores, for example, may be indicative of poor effort, a child being over extended with outside activities, etc. This may lead to conversations with parents to elicit more information about work habits, interests, and time management that enlists the parents as partners in the teaching and learning process.

It is important to recognize that reasoning skills, like achievement, can be taught and learned. In other words, reasoning is not an innate, immutable characteristic, but rather

something children may learn and develop. And while reasoning scores should be related to achievement scores, they do not predict them.

Q: How can we use the item analysis results as a helpful tool?

A: The Item Analysis is one of the most useful reports for teachers. To use it to best advantage, teachers should examine important data points. For example, consider how individual and groups of students performed on areas within the content, e.g., *explicit information, inference, analysis* as parts of the larger area of Reading Comprehension. Next, look for students whose scores are 15-20 points lower than the norm group scores you are comparing them to, and investigate reasons for that. This will give you an idea of where within the content area student(s) may need more help or might benefit from enrichment if their scores are well above the norm group.

It is also important to examine the percent correct of items for the norm group to which you are comparing your students. A low percent correct for an item is an indicator that it is a difficult question designed to discriminate at the top of the achievement level. Students who get many of these items correct may be good candidates for enrichment.

If students are showing weakness in the same content and areas within that content over successive years, that information may shine light on an area of the curriculum that needs review or development. Before rushing to judgment, however, you may want to see if the student did not finish an area of the test. In such cases, it is a good idea to see what percent of the questions the student answered correctly among those reached. This will provide better information about content mastery and may lead to helpful conversations with students about why they did not finish and work on study skills as a goal. Once content items and areas of the curriculum have been examined in this way and improvements are made in certain areas, staff development on implementation may follow.

Q: How can we help parents interpret test results and understand norm group comparisons?

A: ERB recommends that you explain what a norm group is and how the CTP 4 is designed to differentiate among students at the highest level. Start with the national norm, which is representative of students in schools across the country, including high and low socio-economic groups, diverse ethnic categories, and various geographic areas. The independent and suburban public school norm groups represent more selective schools and learners. Therefore, comparison of students to either of those norm groups typically represents a more similar group of learners, making it more difficult to outperform others in the group. That is why the CTP 4 is viewed as a "reality check" on curriculum and serves to help schools identify how well their highest achieving students compare to similar achievers elsewhere, as well as how their average and lower achieving students compare to independent and suburban public students in those respective categories.

Many of the students who take the CTP can be thought of as scoring in the top 20% of students in the country (meaning the 80th percentile or above). When you compare these students' scores to the national norms, the students are going to rank very high. But when you compare your students' scores to the independent or suburban public

school norms, you are comparing the students within the top 20% against each other. A particular student may be doing very well in school, do well on the test, and rank in the 65th percentile among independent or suburban public school students. This is a result of the caliber of the norm group.

When making norm group comparisons, schools should also be careful to consider the number of students in each grade at their school. Norm group comparisons for groups with **fewer than 30 students** on a grade **are less reliable** given the impact each learner has on the average for the group. Outliers in such small groups can similarly impact greatly on group averages.

It is important to remind parents that the CTP is only one piece of the puzzle. Class work, homework, and teacher observations provide valuable insights. In fact, the CTP may confirm a strength or weakness that a teacher has already discovered. The test is good for pinpointing unknown strengths and for catching students' needs as soon as possible so appropriate intervention strategies can be implemented.

Q: What are percentiles and stanines?

A: A percentile is the percent of students in a norm group scoring at or below a particular score. It should not be confused with percent correct. It indicates a student's ranking within the norming sample. For example, a student scoring at the 79th percentile in reading comprehension in the independent norm group scored better than 79 percent of the students in the norm group taking that test.

Stanines, like percentiles, represent a student's relative standing in a reference group. It is the division of a normal distribution of scores into 9 equal parts ranging from a low of 1 to a high of 9. Typically, scores of 1-3 are considered below average, 4-6 average, and 7-9 above average. ERB recommends a difference of two stanines for deciding if there is a significant difference between scores, whether viewing content areas within the same student's scores, comparing students to others in the class or a norm group, or comparing scores of the same student over three or more years.

Q: Are workshops available for teachers on ERB tests?

A: ERB has consultants who are available for workshops and presentations. Contact information for them is available on the ERB Web site (www.erbtest.org) under the CTP 4 or WrAP link and then "Contact ERB."

Q: Is there a help line available from ERB?

A: Member schools may contact Sid Barish at ERB by phone [(800) 989-3721 x 308] or e-mail (sbarish@erbtest.org) at any time if they need assistance with questions about test interpretation. In addition, test consultants are available to come to schools or conduct regional workshops for a consortium of schools who want help with interpretation and use of test results.

Q: Is the CTP 4 useful as an admission test?

A: All of our materials specify that the CTP 4 is designed for use as a standardized achievement test to inform curriculum and instruction and **not** as an admission test to predict success.

Q: What help can ERB give in easing test anxiety?

A: One of the best ways to help students relax when taking the CTP 4 is to maintain its status as a low-stakes test. Its purpose is to help pinpoint specific areas of instruction where special assistance may be needed or to reveal exceptional abilities that deserve recognition. When used in this way, the CTP 4 test results can help teachers determine where classroom instruction needs to be adjusted to meet the needs of students. As students and parents begin to see meaningful attention being paid to test results, anxiety is often overcome by a sense of wellbeing. In many cases, how test results are used determines how students feel leading up to the tests. Testing in the fall may also reduce both parent and students anxieties, because it is more apt to be viewed and used to inform instruction, curriculum, and staff development decisions.

Q: What is the best way to view CTP 4 trends?

A: The scores on CTP 4 are more than accurate enough to be useful, but no test measures perfectly. Test scores are subject to error . . . in part because many factors may affect a student's test performance. Moreover, test questions are merely samples of all the questions that might be asked about a particular subject. Therefore, the greater the body of work that is being assessed, the more accurate the assessment will be.

Accordingly, several consecutive years of test results usually provide a better estimate of a student's knowledge and abilities than scores obtained in a single year. While one year's results may suggest things worth investigating, they will not necessarily tell the whole story. That is why ERB recommends that test results be considered in combination with teacher appraisals of other work being done by students in the course of the school year. Only when information from a variety of sources is assembled can one begin to appreciate the myriad abilities, traits, and qualities each child has.

The foregoing may be reproduced without permission from ERB if the content is not altered.
Revised October 2007

Useful Terms for Understanding Test Results

Percentile Rank—Percentile rank indicates a student's standing in relation to the rest of the norming sample. They are not the same as percent correct. A percentile rank of 79 indicates that the student scored higher than 79 percent of the students at the same grade level in the norming sample being compared. If a student maintains his or her percentile rank from one year to the next it means that the student's achievement has grown about as much as other students in that grade over the same period of time. Relatively consistent percentile ranks over time indicate that growth has been about what would be expected from one testing period to the next.

Raw Score—the count of items correct or points given

Scale Score—a conversion of raw score to a standardized score in order to make it possible to compare results with other students in the chosen norm group and to allow tracking of student performance over time.

Stanines—indicates a student's relative standing in a reference group in a range from 1 (lowest) to 9 (highest) and shows a single-digit measure of where a student is performing. The table below shows the relationship of percentile ranks, stanines, and percent of students in each stanine.

Percentile Rank	Stanines	Percent	
96-99	9	4	
89-95	8	7	Above
77-88	7	12	Average
60-76	6	17	
40-59	5	20	Average
23-39	4	17	
11-22	3	12	
4-10	2	7	Below
1-3	1	4	Average

Median—the middle score in a group (half scored above and half below)
Example: 7 in class with scores 25,30,40,45,60,65,70 Median=45

Mean—the arithmetic average of all scores
Example: 25,30,40,45,60,65,70 Mean=47.8

Normative Data (norms)—a norm group is a group of test takers whose scores are used as a basis of comparison. Norms are statistics that describe the performance of the norm group. Here are examples of three ERB norm group comparisons.

National Norm—compares the performance to that of the group that represents students at urban, rural, suburban, large, small, and high and low socio-economic ranges;

Independent Norm—compares performance to that of students attending independent schools that are members of ERB;

Suburban Norm—compares performance to that of students attending select public schools that are members of ERB.