Executive Summary

It is imperative, as the educational trends and legislative reforms continue to target a strong relationship between scientifically based data and instruction, that educators have innovative ways to assess students which provide both diagnostic and growth data. Computer adaptive technology, such as Scantron’s Performance Series℠, is an efficient and effective way to assess student performance and progress and immediately access assessment driven, standards-based objectives that can be applied to classroom instruction.
Traditional standardized paper-and-pencil tests are inadequate in delivering tests that are adaptable, based on state grade level expectations and individual student performance. They also fail to provide immediate results, and the delayed data from standardized tests are disconnected with the curriculum and do not provide information that is useful in diagnosing strengths and weaknesses of students (by individual, school, and district) on a consistent scale across all grade levels. This paper identifies the gaps in traditional standardized assessments, describes the history and development of computer adaptive tests, highlights the benefits of adaptive testing, and discusses the applications of data from computer adaptive tests.

Identifying the Gaps in Traditional Standardized Assessments

It's an all too common scenario: students spend days, sometimes weeks, taking standardized assessments in the Spring; advanced students are bored with the tests, while struggling students are frustrated and feel an impending sense of failure; the tests are sent away to be scored; the results are sent back to the school, where they are eventually shared with the teachers of those students. Unfortunately, by the time the test scores are available, it is the fall or even the winter of the following school year, and the teacher doesn’t even have those students in his/her class any longer. Little information is gleaned from the data about student performance above and beyond the grade level averages. Further, the data for those students, whose performance was in the center of the targeted performance curve for their grade level, is now irrelevant, because they are now in a new grade level, with new teachers, who have launched into a curriculum with a new set of instructional objectives. As instruction unfolds, there is no way to use the data to measure student progress.

Delayed Feedback
The lack of immediacy in typical assessment feedback renders it useless. Typically, annual standardized assessments are given at the end of a school year, and the results are not available until the following year, while months of instruction have occurred since the assessment was administered. So, the data is not immediate enough to be diagnostic, it is rarely communicated with the teachers who actually teach the students by the time the data is available, and the data gives no information beyond grade level skills, which are no longer taught. At a recent U.S. Department of Education Secretary’s No Child Left Behind Leadership Summit, Glynn Ligon, Ph.D., asserted, “The lag time between testing and availability of the data limits the benefits of assessments and is an Achilles heel for assessments and No Child Left Behind” (Ligon, 2005).

Testing Biases
Traditional paper-and-pencil standardized assessments are also increasingly under fire because of the links between student scores, socioeconomic status, and ethnicity. Standardized achievement tests are designed to yield comparative scores that show who outperforms whom. This means that to secure the degree of score spread among test takers needed to make such comparisons, the national or state tests contain many items directly linked to students’ socioeconomic status. The test results morph into a measure of the composition of a school’s student body, rather than a measure of true student performance (Popham, 2005).

Weaknesses in the Measurement of Individual Achievement and Progress
Current federal legislation, such as the Individuals with Disabilities Education Act (2001) requires the reporting to parents of individual student progress, and proposed bills to reform No Child Left Behind share a common thread: the measurement and tracking of individual student achievement and progress (Senate Bill S. 24 & House Bills H.R. 224, 1506). Most standardized tests, however, do not report standards based scores and scaled scores across all grade levels, and are therefore unable to track student achievement and progress longitudinally. They are also often limited to annual administrations, which prevents the analysis of mid-year student growth, or gains, as well.
Results that Lack Specific Instructional Applications

Typically, the results of criterion based assessments provide student score information, but rarely provide timely, instructionally meaningful feedback to students or teachers (Popham, 2005). The data is one-dimensional and lacks any kind of relevant instructional applications, especially with broad ranges in student scores. Scores for students on either end of the spectrum fail to give any information that lends to an application of the data in the instructional decision making process. How does a percentile ranking alone, inform an educator about appropriate instructional objectives for a student? It doesn’t. What does a grade equivalency, as an average, tell teachers about what students have mastered in individual skill areas? Very little. What does a simple percent correct score tell about the highest item difficulty level that a student is able to reach independently? Nothing at all. Teachers are becoming increasingly frustrated with sacrificing precious instructional time on tests that provide little, if any, information about students that can be immediately used to inform classroom instruction. According to research by the Mass Networks Education Partnership, teachers fundamentally think in terms of the needs of individual students, and want technology to help them with the diagnosis of students’ unique learning needs.

Building Confidence in Computer Adaptive Technology (CAT)

Assessment measures that utilize computer adaptive technology have been used with confidence across industries, with a wide range of examinees, for a variety of purposes. A well designed CAT test is valid and reliable, with content that is constantly reviewed, updated, and analyzed.

The History of CAT

In 1905, Alfred Binet achieved a major advance in the area of intelligence testing. He asserted that with the diagnosis of the individual, rather than the group, there was no reason that everyone had to take the same test in order for it to be fair. By ranking the items by difficulty level, he administered test items that either became easier or harder, based on the person’s previous responses. Although his calculations were somewhat rudimentary, they became easy to implement with the advances in computer technology. By 1968, the Item Response Theory Model was developed (Lord & Novick, 1968), and was further detailed in the Rasch Model (Wright & Stone, 1979, 2004).

Computer adaptive tests are increasingly being used by government, professional, corporate, and educational organizations. The U.S. federal government has long been a champion of computer adaptive testing. In the 1970s and 1980s, the Office of Naval Research became involved in the research and development of CAT, and the Education Department, a decade later, paid for new computer adaptive reading tests in foreign languages. Even then, department officials praised the method's potential for school-based testing. The military uses CAT assessments to assess a number of skills and aptitudes, using tests like the Armed Services Vocational Aptitude Battery. Admissions/entrance exams such as the Test Of English as a Foreign Language (TOEFL), Graduate Management Admission Test (GMAT), Graduate Record Exam (GRE) are all delivered through CAT tests. Professional certification tests have also been administered using CAT tests, such as the U.S. Patent and Trademark Office Patent Bar Exam and the American Association of Clinical Pathologists Certification Exam. More and more businesses and corporations are also using CAT tests. Google, for example, uses computer adaptive tests in their hiring process, as an efficient way to find potential employees with the necessary skill sets for job success. Fifteen states offer computer-based assessments delivered via the Internet, and another five have piloted computer-based state assessments. Of those fifteen states, six use computer adaptive tests, and one is piloting computer adaptive tests (Education Week, 2005). Many states are looking at computer-based testing because it provides immediate feedback (Branigan, 2004).

Performance Series™, powered by Scantron, a leader in data collection for over thirty years, is an Internet-delivered, standards-based, computer adaptive assessment that provides valid and reliable diagnostic assessment data. Each assessment is adapted for each student, is aligned to individual state standards, and links to instructional applications that can help educators design formative assessment-based instruction.

Applying Adaptive Technology to Diagnose Student Performance and Progress Copyright © 2005 Scantron Corporation. All Rights Reserved.
Confidence in the Validity and Reliability of Performance Series℠

Confidence in the scores generated by any computer adaptive test rests on its accumulated evidence of validity. Scantron conducts Content Validity studies in two areas: Item Validity, or the degree to which the test items are relevant in measuring the desired content area, and Sampling Validity, or the degree to which an assessment includes items that span a given content area. A large component of the confidence in computer adaptive tests lies in the confidence in the items, themselves. Developing the item bank for Performance Series℠ was an intensive and comprehensive effort by a large team of item developers, teachers, and educational consultants. To ensure the highest level of quality, all items are developed by trained specialists with extensive background and experience in education. Once items are created, they are submitted to the item development team at-large for reviews of grade-level and contextual appropriateness, and are returned, if necessary, to the writing team for editing and resubmission. Items are also reviewed by the editor team, which consists of professional educators (credentialed teachers and university professors) from around the United States and Canada. This team is tasked with carefully analyzing the content of each item, examining response choice construction, and proofreading all questions and answers. This review includes areas such as age appropriateness, interest level, bias, sentence structure, vocabulary, clarity, grammar, and spelling. A special team of consulting educational experts, from a sample of national educational communities, representing diverse cultural backgrounds, also reviews and analyzes all item content.

Another way to build confidence in the test items is to avoid the risk of their overexposure. Scantron introduces an average of 5,000 new items for all four subject areas every year through an online calibration process known as item embedding. Students who take Performance Series℠ assessments see embedded trial items, which are analyzed to determine the item’s performance statistics. These items have no effect on the student testing experience or scores, and provide an efficient way of frequently replenishing the item pools.

To demonstrate the concepts of item and sampling validity of Performance Series℠ in a more quantitative manner, Scantron has also examined the Inter-Testlet Correlation between the component testlets, or units, within each content area. In a 2004 study, the Mathematics inter-testlet correlation ranged from 0.512 to 0.876. The correlations between the individual units to the overall score ranged from 0.741 to 0.876. Reading inter-testlet correlations ranged from 0.744 to 0.967. Correlations between the individual units and the overall score ranged from 0.823 to 0.967. Similar results were found in Language Arts and Science.

Scantron has also been engaged in Concurrent Validity research, which is a form of Criterion-Related Validity, since the initial release of Performance Series℠. Correlation between student scores on Performance Series℠ and standardized assessments have been conducted on the following: Stanford Achievement Test (SAT) 9, Iowa Test of Basic Skills (ITBS), the California Achievement Test (CAT) 6, California Standards Test, and the California High School Exit Exam. A high predictive correlation between Performance Series℠ and the ITBS, and Performance Series℠ and the Dakota State Test of Educational Progress (STEP), have also been measured.

It is a common misconception by teachers that they cannot have confidence in comparing student test scores unless the tests are identical and uniform, and this perception has been a roadblock in the use of individualized, adaptive content (Shepard, 2000). However, a necessary condition in computer adaptive tests is Unidimensionality, which means that the items in the pool all measure the same construct. Dimensionality analysis of Performance Series℠ items is therefore a part of the item analysis. Adaptively administered items produce tests with equivalent measures and produce parallel tests, even though the tests do not have identical questions (Steinberg et al, 2000). Unidimensionality makes it possible to rank order the items by difficulty, allowing inferences on the items to be made based on the students’ measured ability level.
Computer adaptive tests have the potential to reach a higher degree of Reliability. This can be achieved by using the Standard Error of Measurement (SEM) as one of the test stopping criteria. An SEM of less than 0.30 logits is roughly equivalent to a conventional reliability coefficient of 0.91. This is the targeted SEM used in Performance Series. Most standardized fixed-form assessments are designed to measure the average ability range for the specific grade being tested. This compromises the precision (reliability of scores) for students far above or below this range. In contrast, computer adaptive tests draw content from a wider range of abilities, making it possible to calculate highly reliable measures.

Recognizing the Benefits of Testing with Computer Adaptive Technology

Computer adaptive tests have the ability to bridge the gap created by standardized assessments, by providing efficient and individualized test administrations that are standards-based and non-biased, which provide immediate results that can be used to show student progress, can serve as the basis for test-based research into best practices, and provide instructional feedback that can be applied in the classroom.

Unique advantages of Internet-based computer adaptive tests, including creating individualized test content that can be delivered in a relatively short amount of time, eliminating the inefficient, paper based “trickle down” method of communicating student assessment data, and easing the transition of students between grade levels by providing student growth data that can be digitally shared between school buildings.

Instructionally Sensitive, Individualized Assessment Content

Traditional fixed form assessments are unable to provide individualized tests for every student in a classroom or school, so a handful of students invariably take a test that far exceeds their ability level, while other students find assessments “too easy”. Tests that are much too easy for test-takers are likely to be a waste of time, and provoke unwanted test-taking behaviors, such as careless mistakes and the deliberate selection of incorrect answers to what are perceived as “trick questions” by highly analytical students. On the other hand, questions that are too hard produce uninformative test results, because test-takers cease to seriously attempt to answer the questions when they sense their own failure, and they resort to guessing (Linacre, 2000). Performance Series creates individualized versions of assessment based on student performance in a variety of units in each content area. All students, even advanced learners, are challenged by Performance Series. Struggling learners, however, are not as easily frustrated by the failure that they have become accustomed to in traditional criterion assessments.

Reading content of Performance Series spans grades 2-10 across four units: Vocabulary, Fiction, Non Fiction, and Long Passage. The Vocabulary section includes objectives that relate to identifying vocabulary in context and in isolation, analogies, and sentence completion. Fiction assessment items include basic learning objectives, such as differentiating between fact and opinion, identifying the main idea, drawing conclusions, identifying story elements (character, setting, plot, etc.). They progress to more advanced objectives, such as identifying contradiction, interpreting the meaning of figurative language, analyzing arguments, identifying the relationship of passages, and inferring character traits, comparing and contrasting implicit elements. Non Fiction skills range from evaluating the author’s purpose, identify details, and identifying the intended audience, to determining the tone of the passage, analyzing the reasoning in the passage, and analyzing arguments in a passage. Long passage skills span basic comprehension of instructions, directions, and tasks, to reading and understanding passages that contain scientific information, making connections between passages, and recognizing the author’s attitude reflected in a passage. Readability of the reading passages and questions are analyzed using computer-based algorithms, including the Spache, Dale-Chall, Flesch-Kincaid, Flesch Reading Ease, Flesch Grade Level, SMOG, Fry Graph, and Powers-Sumner-Kearl. A Reading Rate is also calculated,
based on a silent reading rate calculated by dividing the total time it takes a student to read each passage into the total number of words in each passage, generating a “words per minute” rate.

**Math** units in Performance Series℠ are *Number & Operations, Algebra, Measurement, Geometry, Data Analysis & Probability*, and cover problem solving and computation skills for grades 2-9. *Number & Operations* skills vary from adding single digits, comparing groups of coins, and connecting fractions to pictures, to performing scalar multiplication with matrices, calculating compound interest, and solving real world problems by applying permutations/combinations. The objectives in the *Algebra* unit extend from completing counting patterns, identifying basic patterns, and writing numbers sentences to multiplying polynomials, solving quadratic equations by graphing, and solving systems of linear equations. The concepts in *Measurement* begin with basic skills such as converting time, measuring weight, and ordering objects by length, and advance to solving problems involving time conversions, calculating volume, and determining arc length. *Geometry* skills range from identifying shapes and figures to recognizing and evaluating tangent, sine, and cosine, solving real world problems using trigonometric concepts, using the Pythagorean Theorem, and translating ordered pairs. From reading basic tables and pictographs to interpreting stem-and-leaf plots, determining the probability of dependent events, and solving real world problems using tree diagrams, the items in the *Data Analysis & Probability* unit also span a wide range of skill levels.

**Language Arts** assessments in Performance Series℠ have content from grades 2 to 8, and include the units of *Capitalization, Parts of Speech, Punctuation, and Sentence Structure*. The scope of skills in the *Capitalization* unit ranges from determining the correct capitalization of cities/states, days/months, and first names to editing to identify the correct capitalization of proper adjectives, determining the correct capitalization of an organization, abbreviation, or acronym, and recognizing the correct capitalization of personal titles. The *Parts of Speech* unit includes using adjectives, adverbs, and nouns in a sentence, and extends to skills like using past progressive, present progressive, and present perfect verbs in a sentence. Encompassing skills from determining the correct placement of commas, apostrophes, and question marks, to identifying the correct use of semicolons between two independent clauses, determining the correct use of single quotation marks within double quotation marks, the *Punctuation* unit covers a range of difficulty levels. The Performance Series℠ Language Arts *Sentence Structure* unit sweeps from skills like determining subject-verb agreement and constructing simple sentences to revising a sentence to correct a run-on and locating a parallel structure error in a sentence.

**Science** units in Performance Series℠ currently include *Living Things, Ecology, and Science Processes*, which encompass skills for grades 2-8. The *Living Skills* unit spans skills like associating animal parents with their offspring, identifying insects, and grouping animals and plants by their similarities/differences, to understanding natural selection, understanding Linnaean classification, and identifying the function of spores and/or plants that produce them. From associating living things to their habitats, adding organisms to a basic food chain, and recognizing pollution, to interpreting a population chart, recognizing examples of symbiosis, and understanding the causes/effects of acid rain, a broad range of skills is covered in the *Ecology* section. In the *Science Processes* unit, skills range from associating a model to what it represents, identifying the senses, making predictions by identifying patterns, to recognizing sources of error, using data from experiments to make generalizations, and making predictions using data from a line graph.

**Individualized Test Administration**
Performance Series℠ only takes an average of 45 minutes, per subject area. The presentation of items is individualized for each student, based on his/her unique responses in the progression of the test. The first item, selected from a large pool of items, presented to a student taking a test, is an item near the difficulty level that matches the student's grade level. The initial question varies from student to student, even for students in the same grade level. Then, each subsequent item is selected based on the student’s performance on previously presented items. Students who demonstrate higher performance have increasingly difficult questions, while students who demonstrate lower performance have less
difficult questions. Using this item selection algorithm to administer test items makes the estimate of student performance more and more precise as the test progresses, by continually presenting test items that hone in on the student’s true achievement level. This adaptive test delivery method means that students each have a unique combination of test items, and that the length of the test varies from student to student.

Tests can further be individualized by giving individual students a Modified Starting Point in each subject area, as a testing accommodation for students with special needs or for English Language learners. The Modified Starting Point allows students who may be performing several grade levels below the grade level they are in to begin at a lower level of item difficulty. Students who teachers believe to be performing above grade level can also have a modified starting point that begins the test with items at a higher level of difficulty. In theory, the starting position should have no effect on the student’s final ability estimate, since the test is adapting to their performance level. In practice, however, a Modified Starting Point helps prevent students who are above or below grade from becoming too easily bored or frustrated with what appears to be an easy or difficult test. Once a student has taken a test, Performance SeriesSM employs an Intelligent Start for future test administrations, which selects items for the start of the test that are appropriate to the student’s ability level, based on his/her performance on the previous test. Performance SeriesSM also archives information about the items each student has seen, ensuring that no item is presented to the student more than once, both on the same and subsequent test administrations.

Seamless Alignment to State Standards
Although standards provide the framework around which curriculum and instruction are built, standardized norm-referenced tests rarely measure and report individual student growth as it relates to state standards. Criterion-based tests can measure how well a student has progressed in specific skill areas based on standards, but traditional fixed-form criterion referenced tests fail to measure student ability when it falls above or below grade level expectations for any particular skill. While no assessment can be an exact measure of student ability, computer adaptive tests, like Performance SeriesSM, provide a more accurate measure of student ability by individualizing the test for each and every student, according to the state standards in his/her state.

In creating the Performance SeriesSM item bank, Scantron Corporation has targeted the need for accurate measurement of state and national standards. To achieve that end, Scantron developed an extensive list of skills that correspond to those critical learning objectives most commonly taught throughout the country. With alignments to national and state standards documents in math, language arts, and reading, and science from across the United States, Performance SeriesSM has a high degree of correlation to state and national standards. Districts/schools can choose to align their online site to their state standards, which filters the items used for the tests in accordance with the state standards as a part of the constrained item selection that ensures that the items selected in an adaptive test are linked to the content standards—a requirement under No Child Left Behind.

The standards alignment of Performance SeriesSM assessments apply not only as a criteria for item selection, but are also incorporated into the reporting and application of the assessment data through standards-aligned suggested learning objectives. Scantron utilizes the most updated standards, and any changes to state standards documents are applied within 6 weeks of the publication of new standards.

Non-Biased Testing
All Performance SeriesSM items are reviewed by bias editors to ensure that the items are unbiased, which follows the Code of Professional Responsibilities in Educational Measurement (1995). Bias editors analyze how many stories/questions contain male or female main characters, and whether these characters use an active or passive voice. In addition, bias editors analyze which stories/questions contain ethnic or cultural diversity. When there is a significant disparity between genders, a lack of cultural diversity, or a misrepresentation of any kind, the item development team makes adjustments accordingly. In addition, the items are analyzed for any differential function that may be present, to
eliminate items for which ability-matched examinees of different subgroups have significantly different probabilities of answering correctly.

While the items are carefully examined to avoid any bias, there is also evidence that the format of computer adaptive tests is less biased than standardized fixed form assessments. Numerous studies have shown that the differences in student performance between different socioeconomic groups and between different ethnic groups are less in computer adaptive tests than in standardized paper-and-pencil tests. A comparative study, for example, of the performance of students on the Standardized Achievement Test (SAT) and a computer adaptive test with an item pool containing items with the same range in difficulty level revealed that while the Verbal and Quantitative scores of Caucasian students remained virtually the same, the scores of African American, Hispanic, and Asian students increased significantly on both tests (Schaeffer, et al, 1998). The differences are attributed to increased student motivation on computer adaptive tests and to the scoring structure of the test, itself: while most paper-and-pencil tests are scored using a method that gives hard and easy items equal weight, a computer adaptive test measures student proficiency by the difficulty of the items that were presented as well as the number correctly answered. So, if a student misses an item because he/she is culturally unfamiliar with any of the content of the question, the adaptive nature of the test will still be able to zero in on the student’s ability level, based on the student’s performance on other items of equal or greater difficulty.

Applying CAT Data to Diagnose Performance and Measure Progress

For assessment data to be useful, it must be immediately available, and it must provide information on a consistent scale that can be applied and interpreted across all grade levels for high level analysis. It must provide specific, standards-based objectives for individuals students and classrooms, and be linked to instructional resources, for it to be useful for classroom applications. The data must be shared easily, so that collaboration in instructional decision-making is possible. And, increasingly important as expectations increase and resources diminish, the data must have the capacity to be easily disaggregated if it is to be used as a form of test-based evidence, so that resources can be directed to programs and strategies that have scientifically based evidence of effectiveness.

The Value of Immediate Feedback

Rather than tying up time and resources sharing information with educators through the typical “trickle down” approach to sharing student data from the top, down, both teachers and administrators have access to student data immediately after any student has completed any test. This “teacher first” reporting allows for teachers to immediately share assessment feedback with students and parents, as well, rather than sharing the data months after the tests were administered. Student data can easily be immediately communicated with parents and students, through composite student profiles, which are automatically created for each student, and include detailed descriptions of each data point. Assessments must be able to provide immediate feedback, so that students can have an active role in their own learning in light of understanding of what it means to improve (Black & William, 1998).

High-Level Analysis of Data

The average classroom has a large breadth of knowledge and ability level, especially in secondary school settings, where the learning gaps have gradually become more and more pronounced. Traditional assessments, however, only assess knowledge within a single grade, which means that using traditional fixed form grade level assessments, only the level of knowledge for students performing at grade level in every unit, across every subject area, can be accurately measured. In Performance Series™, scores are reported as Scaled Scores, on a scale from 1300 to 3700, so that student performance and progress can be measured and compared, across grade levels. Ability estimates for each student are also reported as SIP (Standards Item Pool) Scores. The item pool score represents the expected score or expected proportion of items correct if the student was to see every item available for his/her grade level, standards-constrained item pool. A National Percentile Rank (NPR) is also available, and compares student Scaled Scores to a normative sample for mathematics and reading Tests. The
NPR of a particular student is the percentage of students whose scores fall below the Scaled Score for the given student.

Performance Series℠ reports are organized into Summary reports, Gains reports, and Percentile reports. The data in the reports can be sorted by column header, graphed by row or column, and exported into spreadsheets. The scope of information available in the reports is based on the user’s rights and the location in which the data is viewed. At the district level, Summary, Gains, and Percentile reports can be viewed for the district as a whole, or can be broken down by the courses in the district, the grade levels in the district, or by groups (user defined subgroups). School level reports, which can be viewed by administrators at the school or by administrators at any level above the school in the site hierarchy, show Summary, Gains, and Percentile reports broken down by School, Staff, Student, Course, Class, and Group. Reports for teachers, which also include the Summary, Gains, and Percentile reports, include only data for students in classes to which teachers are assigned, and can be broken down by Student, Course, Class, and Group. Testing status reports, including Unfinished and Active Tests, Checklists of Students Tested (sorted by week, testing period, and year), and Spoiled Tests are also available to school administrative users, so that teachers, administrators, and test proctors can manage the school-wide testing process and ensure that testing data is captured for every student.

**Summary Reports** show the total Student Count, Mean Scaled Score (SS), Standard Error (SE) of the Scaled Score, and the overall Standards Item Pool (SIP) score for each subject area test. Summary reports are also available for each subject area test, and include the Student Count, mean Scaled Score (SS), Standard Error (SE) of the mean Scaled Score for all subjects, and the SIP score for each unit within the subject area.

**Gains Reports** show the differences in scores between testing periods. Performance Series℠ can be delivered an unlimited number of times per year, and Gains reports can be viewed for tests that are delivered in testing periods that are at least 12 weeks apart. Gains reports are available for each subject area, both for the individual and group level. Group Gains reports list the student count and the mean Scaled Scores (SS) and Standard Error (SE) for Testing Periods 1 and 2, with a calculation of the mean difference and SE of the mean difference. If the difference is a positive value, it is indicated as a gain in green. If it is a negative value, it is indicated as a loss in red. If a difference is not statistically significant, it is indicated with a footnote.

**Percentile Reports** show the National Percentile Rankings (NPR), based on the National fall/spring norms of students at the same grade level. Over 70,000 students are included in each of the fall and spring norm groups, and the groups are representative samples of the gender/ethnicity composition of the nation, as a whole, across four geographic regions.

The Application of Data in the Classroom

It’s no wonder that teachers balk at losing classroom time to deliver assessment, when traditional assessment data provides little information that can be applied in the classroom. Providing scores and rankings, although it serves a necessary purpose for high level analysis and reporting, is not enough, by itself, to provide comprehensive feedback. Instructional applications of the information must be utilized for the information to truly be considered feedback (Black & William, 1998). With Performance Series℠, teachers are able to actually conserve instructional time by targeting the instruction, and not wasting time on skills that students and classes have already mastered. Instructional applications are inherent in Performance Series℠, through the automatic creation of individualized Suggested Learning Objectives, which are aligned to state and/or national standards. Unique to Performance Series℠, among other kinds of computer adaptive tests, the Suggested Learning Objectives are available for each student, in each unit, in every subject area tested. The learning objectives provide more relevant information for classroom applications of student data than score reports, alone. Each Suggested Learning Objective specifically references the aligned state standards (if districts so choose), in a checklist of up to 10
suggested learning objectives for future instruction. Checklists of up to 10 objectives that a student has attained in each unit are also available.

Many schools use these learning objectives as a way to incorporate standards-based, measurable, goals for the Individual Education Plans (IEPs) of students with special needs. The use of Suggested Learning Objectives is not limited to application by educators. They also provide information to students that can help them identify their own learning goals as they measure the progress in their skill attainment over time. Effective assessment must provide feedback to students that emphasizes the strengths and weaknesses of each student’s performance (Gronlund, 2002). By including both the Successfully Attained skills and the Suggested Learning Objectives, detailed feedback can be shared with each student that focuses on successful elements of his/her performance as well as the areas of weakness.

Even further, the Suggested Learning Objectives are linked to Scantron’s Skills Connection resources, which allow teachers to create individualized, editable study guides and assessments, such as pre-tests and post-tests, which can be used to assess and monitor student growth on the specific objectives. Scantron also sponsors a free website at http://www.statestandards.com, which provides links to state-standard/skill specific web-based resources.

While teachers can view the individual Suggested Learning Objectives for each student, they can also view profiles of learning objectives, by class, with specific information about which students have mastered which standards-aligned objectives. Class Profiles are arranged by subject area, and they list in order, the objectives completed by the largest number of students to the objectives completed by the fewest students. A detailed checklist is available for each objective that lists the names of students who have and have not completed the objective.

The Communication of Data for Collaborative Instructional Decision Making

Teachers need to be able to reflect on assessment data and consult with other teachers about the instructional changes they would like to make (NRC, 2001). Attempting to communicate paper based assessment data, however, between educators within and between schools is extremely tedious. If assessment data is to be intelligently shared, it is imperative that data tools, including assessment management systems, be able to communicate with one another (Bernhardt, 2005). It must also be easy to manipulate and organize. With the average secondary teacher having 167 students on his/her roll (NCES 2005025), robust paper assessment data is simply unmanageable for collaborative applications. Making data available to every teacher who works with every student is almost impossible in a paper based assessment data system. With an increase in student support services, such as tutors, specialized teachers, and counselors, especially for learners with special needs, the ability to communicate data for instructional collaboration is imperative. By providing assessment results electronically to all staff who work with any particular student, as soon as that student has completed a Performance Series™ test, educators and support staff can intelligently access and interpret data for all of the their students, across subject areas, and focus on ways to collaboratively help the student meet his/her instructional goals.

Whether the organization’s philosophy is site-based administration or centralized control, Performance Series™ can accommodate almost any organizational structure and assigns unique site identifiers to each physical/virtual location with up to five sub-levels beneath a “top level” site. Each level has access to view the information at and below that site, allowing for a customized administrative site structure and controlled access. The type of data that is accessible to each staff member can be specifically defined through a variety of user rights. While administrative data may be imported via .csv spreadsheets into Performance Series™ from other electronic databases, the assessment results from Performance Series™ may also be exported through simple .xml spreadsheets for integration back into a school or district information system and/or data warehouse. Data can also automatically be communicated between locations under a top level site (such as a district or state) through the Rollover of student data from one year to the next. The data “follows the student”, regardless of the changes in his/her school or
class enrollment, which gives access to a history of the student’s assessment data to any teacher or support staff member that has the student assigned to his/her class.

The collaboration possible through the electronic access to diagnostic assessment data also extends past the school walls, and helps facilitate the transition of students from lower grade school buildings to higher grade school buildings. Transitions present a challenge for both students and teachers: students must become familiar with new settings, new curriculum materials, and new teaching styles; educators must familiarize themselves with new students, often with little information or data to prescriptively plan for the curriculum and instruction of individual students. While the constraints of traditional paper based assessment reporting mean that data is rarely communicated between school buildings, teachers at different schools through a district can have access to Performance SeriesSM data, which can be crucial in informing those transitional periods. The data in Performance Series can be rolled over from year to year; building to building, and that data can be intelligently and efficiently communicated for collaborative applications across a district or state. This information helps to ease the student transitions, while providing an efficient means for educators to effectively make data-driven decisions even before their ‘new’ students walk through the doors on the first day of school.

The communication of data should not compromise security of sensitive information, however. Performance SeriesSM incorporates collaborative features into the access of data while also carefully guarding student information. In accordance with the Family Educational Rights and Privacy Act (FERPA), security measures within Performance SeriesSM protect the privacy of student assessment records. All of Scantron’s networks employ multiple security measures including secure passwords, address restrictions, firewalls, and physical security. All administrative tasks, including viewing scores and managing student information, use 128-bit SSL encryption techniques. Schools can also configure on-site security settings to limit access to only certain computers, certain times of day, or certain days of the week, regardless of access codes, which provides a level of intelligence that most schools simply can’t provide through human monitoring of paper data in a classroom or school filing cabinet.

The Application of Test-Based Evidence of Student Achievement

Every school in the country is challenged by the push to show evidence of student growth with limited resources. As time, money, and energy become increasingly unavailable, it is paramount that schools have tools to evaluate best practices and programs as they make decisions about how to invest limited resources. The National Educational Technology Plan recommends that states, districts, and schools “use data from both administrative and instructional systems to understand relationships between decisions, allocation of resources and student resources.” (NETP, 2005) Gathering test-based evidence of student achievement through cross-tabulation of Performance SeriesSM data and site based studies of the factors that affect achievement can help educators determine best practices and evaluate the efficacy of programs. Through Performance SeriesSM, students may be enrolled in user-defined groups that can be used to evaluate the relationship between student achievement and specific conditions that may positively or negatively affect academic progress, such as their participation in academic programs, extra-curricular activities, and intervention programs. If a school is considering eliminating arts programs, for example, what better evidence is there to support arts programs as a means of increasing student achievement than assessment data that shows the growth of students in arts programs surpassing average student progress? If a school considers adopting a new textbook series or curriculum program, what better way to evaluate its effectiveness than by piloting the program and tracking the assessment progress of the pilot program students compared with the rest of the student population through Performance SeriesSM group disaggregation? There are a number of schools that use Performance SeriesSM to gather this kind of research to become better decision makers and more savvy consumers of products and programs in their schools. And, as funding becomes more and more dependent on proof of effectiveness and scientifically based programs in schools, the use of Performance SeriesSM to evaluate factors that affect student achievement will continue to grow.
Conclusion
Standardized tests lack the ability to provide individualized test sessions that adapt to students' unique strengths and weakness. They fail to provide the data immediately, and the data lacks the components necessary for high level analysis above and beyond single grade level analysis—across all grade levels; across every subject area; across every school and district. Traditional results do not provide the meaningful feedback that is necessary for teachers to assess how students are learning and progressing, and the time invested in the actual assessment is wasted without true feedback that can be used to inform instruction. With computer adaptive tests, like Scantron’s Performance Series℠, teachers and administrators can have confidence in a powerful technology that can diagnose student performance and progress on a consistent, valid, and reliable scale.
References


