Perspective™

Integrated Assessment and Instructional Resources System
A white paper from the Assessment & Information group of Pearson

Jason L. Meyers, David Shin, Paul Nichols

March 2008

wp0801
Introduction

Educators face many challenges in their quest to meet the NCLB-stated (NCLB, Public Law 107-110) goal of having all students demonstrate proficiency in the content areas of mathematics, reading, science, and social studies. One way to improve student achievement and meet these goals is through the use of assessments that provide formative test score information. Formative test score information is that which is used to narrow the gap between students’ current state of achievement and the targeted state of achievement.

The concept of formative test score information developed out of Scriven’s (1967) early work in the field of program evaluation. Formative evaluation was intended to foster development and improvement within an ongoing activity while “summative evaluation” was used to determine if the program’s results matched its stated goals. Bloom later incorporated this concept into his Mastery Learning Model (Bloom, Hastings, & Madaus, 1971) in which students do not progress to the next learning objective until they have mastered the current one. In this technique, assessments are embedded within daily activity and the teacher utilizes diagnostic information gleaned from the assessment to implement corrective action by targeting instruction that is specific to the student’s weaknesses.

In the classroom, teachers frequently use assessments (homework assignments, quizzes, tests, etc.) to serve formative purposes in the way that Bloom described. These assessments evaluate student performance and the results are intended to be used to improve and target teaching to the specific
student needs. The goal is to identify and narrow any gaps that might exist between what a student knows and what they are expected to know. In essence, assessment and instruction are inextricably linked.

Many types of assessments can provide formative information. The assessment need not be teacher created nor embedded within daily classroom instruction to be of diagnostic value. Wiliam and Black (1996) state that “in order to serve a formative function, an assessment must yield evidence that, with appropriate construct-referenced interpretations, indicates the existence of a gap between actual and desired levels of performance, and suggests actions that are in fact successful in closing the gap.” Many assessments, including large scale standardized assessments, can provide formative information.

Teachers sometimes need support to use test results to inform instruction. Consider two classmates, Timothy and Angela, who have recently taken the NCLB- required mathematics assessment in their state. Upon receipt of the test results, it is clear that both students failed to reach the Proficient achievement level set forth by the state. In addition to managing the needs of all the other students in the class, the teacher must try to determine which specific areas Timothy and Angela are struggling with and identify and locate student resources to get these two students to where they need to be. When remedial materials are located, how will the teacher know that they match the content on the statewide test? How will the teacher know when Timothy and Angela have mastered the content they struggled with on the test? How can
the teacher direct the concerned parents of these two students towards appropriate materials?

Test results from high-stakes statewide assessments are typically used for accountability purposes, but are less often used for formative purposes. Huff and Goodman (2007) found that of educators that receive state-mandated assessment results, 31% use the results daily to inform instruction, 14% use results a few times a week, 15% use results a few times a month, 21% use results a few times a year, 14% use results only once a year and 7% never use results from statewide assessments to inform teaching. While NCLB requires that assessments report individual student results in ways that allow parents, teachers and principals to understand and address the specific academic needs of students (NCLB, Public Law 107-110), many current assessments do not deliver information about student learning and knowledge to teachers in ways they can use to improve achievement (National Research Council, 2006). Although most student score reports compare student performance to a set of standards and may include strengths and weaknesses in some number of content sub-domains, there is typically no link to additional resource materials that may be accessed to address the observed student weaknesses. So, while large scale assessments can serve as a powerful vehicle for diagnosing and addressing student learning deficiencies, that potential is rarely realized.

With high numbers of students per classroom and the diverse needs of each student, it is often difficult for educators to locate additional resource materials and provide remedial instruction to students who need it while not halting the progress of the more advanced students. This is particularly true
when attempting to utilize score information from assessments that were created outside of the classroom. The need for an integrated system of assessment and instruction is what prompted Pearson to develop the Perspective™ Reporting System and the Learning Locator™ service. These products are intended to realize the full potential of using information from large scale statewide assessments for formative purposes.
Perspective Development

Before unveiling Perspective, Pearson conducted extensive research on what features of an integrated system would be most beneficial to teachers, parents, and students. A staff of usability experts facilitated numerous face-to-face focus group meetings and conducted phone interviews with many teachers, principals, test coordinators and parents. Perspective’s development process incorporated elements of user-centered design (Norman, 1986) that support content validity claims about the system. Just as validity claims about an assessment can be supported based upon the strengths of the test development process, employing this user-centered design method where the users (teachers, students, and parents) are included in the development from the ground up provides strong validity evidence.

Pearson’s research identified three features as being most important to improving student achievement: rapid delivery, audience-specific reports, and easy access to appropriate educational resources. Perspective was developed using these three features as a foundation. As such, Perspective provides rapidly delivered paper and/or online reports tailored to the specific needs of administrators, educators and parents. In addition, Perspective supplies separate web portals for families and educators that take student assessment results and map them to customized resource and intervention materials.
Perspective System

Similar to the work done by Pellegrino & Chudowsky (2003) representing assessment as a triangle composed of the three pillars of cognition, observation and interpretation, Perspective can be conceptualized as part of a larger system of integrated elements. Such a system is depicted in Figure 1. The elements of this system are 1) the assessments that evaluate the degree to which students have learned required concepts 2) the instruction of these concepts, which can occur either within or outside of the classroom, and 3) the student learning of these concepts, which can also occur within or outside of the classroom. Perspective operates at the intersection of these three elements and serves by integrating assessment results with instructional prescriptions to improve student learning.
Figure 1: Components of Pearson Perspective Integrated Assessment and Instructional Resources System.

**Perspective Components**

The Perspective system consists of four components: score reports, an online portal, instructional resources, and a Learning Locator™ service. These components are designed to work together, along with the assessment results from state tests or PASeries, to provide educators, parents, and students with the information they need to narrow the gap between a student’s current achievement level and his or her targeted achievement level. In this section, each of these components is described in more detail. After each component has been described, the way the components work together within Perspective is described.

**Customized Reports**

Perspective customized score reports are presented in three formats: Individual Student Reports for families, Classroom Reports for educators, and School Reports for administrators. All three reports aim to identify areas that require additional attention, areas that have been mastered, and to provide what additional help a student might need to succeed. Printed reports include a targeted Interpretive Guide, which explains the assessment results so the audiences can focus on improving instruction and increasing achievement. Perspective customized reports present a performance summary, performance by content area, performance by demographic group, item analyses, next steps, and a proprietary code, called the Learning Locator, which takes a student’s test performance and links that information with specifically targeted
instructional materials and resources. More information on the Learning Locator is presented in a later section.

**Online Portal**

Perspective’s online reporting and resource portal was designed to provide families and teachers with a collection of tools that allow them to utilize students’ assessment results. The Perspective web portal was designed so that it would be easy to use for students, families, and teachers. Pearson’s usability staff conducts ongoing research to evaluate the extent to which families, educators, and students are able to navigate the system and to determine whether it contains the desired functionality. There is one web portal for educators and one web portal for families.

The Educator website was designed to provide assessment results at the classroom level. Instruction can be targeted to specific student needs in two ways through use of the website. First, at the beginning of a school year or new semester, teachers can view rosters of incoming students and their collective and individual performance levels. A feature called dynamic rostering allows educators to conveniently access a history of their students’ most recent large-scale test results in order to tailor instruction to the individual needs of the students prior to an annual assessment. In addition, after an assessment administration, test results are associated with additional remediation materials to help correct student learning deficiencies. This process of mapping results to learning materials saves time and cuts out the guesswork for the teacher and parents. Resources are provided online,
allowing the teacher and student to promptly address any specific learning deficiencies. Teaching resources, such as unit plans, worksheet answer keys, released tests, and professional development material are also provided on the Educator website.

The goal of the family website is to allow family members to actively engage in their child’s education. It provides resources that are intended to help family members better understand their child’s assessment scores and to locate appropriate learning activities aimed at improving those scores.

**Learning Resources**
Learning resources consist of: activities, instructional materials, quizzes, games, videos, and other materials that have been designed for specific grade levels, subjects, and learning objectives. Within the Perspective website, students can access these learning resources directly or by navigating the site under the teacher’s direction. Students can also browse the library of available materials and work at their own pace.

By using the Perspective system, students can access a personalized set of supplemental resources designed to provide practice on specific learning standards and to encourage students to meet achievement criteria for their existing achievement level and to progress to the next achievement level where appropriate.
**Learning Locator**

A core component of the Perspective integrated system is the Learning Locator — a tool that evaluates student performance by content area and links it to appropriate learning resources. The Learning Locator is the mechanism that drives the integration of assessment, instruction, and learning as depicted in the figure on page 6. A Learning Locator code is printed on each student’s test report. That code can then be used to access learning resources. The goal of Learning Locator is to provide teachers, families, and students with resources that are specifically mapped to the student’s performance on the measured standards.
Learning Locator
The Learning Locator service works through invocation of a four-step process. Those steps are 1) estimation of student subscale score, 2) estimation of a subscale cut score, 3) generation of a student learning profile, and 4) creation of a link to learning resources. Each step employs research-based methods. A detailed list of the Learning Locator features can be found in Appendix A of this document.

**Step 1: Estimation of Student Subscale Scores**
In order to appropriately serve diagnostic purposes, an assessment must yield reliable scores not only at an overall test level but at the level of particular subsets or clusters of items as well. Under Perspective, when a student’s test responses are scored, a statistical analysis program estimates his or her “subscale”, or content area, performance and generates a “learning profile.” Scores from these item subsets, referred to here as *subscale scores*, might represent student performance on learning objectives, subtests or learning standards. To provide valuable formative information, scores are needed that focus as narrowly as possible on the content areas in which a student may be having difficulty (Wainer, Vevea, Camacho, Reeve, Rosa, Nelson, Swygert, & Thissen, 2000). Subscale scores are often of interest to teachers and parents in evaluating their students’ specific strengths and weaknesses (Yen, 1997; Wainer et al., 2000). A subscale score is a crucial component of any assessment that provides formative, diagnostic test score information.
When a test has a large number of items measuring an objective on a test, estimating a score on that objective can often been done precisely and reliably. However, in practice, tests often contain a less than optimal number of items per objective (Pommerich, Nicewander, and Hanson, 1999). Because of this reality, a number of studies have been conducted that investigate various methods for estimating subscale scores (Yen 1987; Yen et al., 1997; Bock, Thissen, and Zimowski, 1997; Pommerich et al., 1999; Wainer et al., 2000; Gessaroli, 2004; Kahraman and Kamata, 2004; Tate, 2004; Shin, Ansley, Tsai, and Mao, 2005). Many of these methods utilize information from performance on other test objectives, or collateral test information, as part of the estimation procedure in order to increase the reliability of scores on the subscale of interest. The precision and reliability of the subscale score ultimately impact how well a student will be matched with resources to address his or her specific learning needs.

The Learning Locator service can implement a variety of subscale estimation methods such as the Item Response Theory Test Characteristic Curve (IRT TCC) and Item Response Theory Expected A Posteriori (IRT EAP) methods. Pearson is currently working to implement the methods of Wainer (Wainer et al., 2000), Yen (Yen, 1987) and Bock (Bock et al, 1997) and that functionality should be available within the next year. More detailed technical information on the estimation procedures and discussion of an empirical investigation of the various procedures is presented in Appendix B.

Pearson works with each client to select the estimation method most appropriate for their needs. Having a toolbox of available methods from which
to choose ensures that each student is matched with the most appropriate learning materials.

**Step 2: Estimation of Subscale Cut Score**

After the subscale score is estimated using one of the above methods, the second step in finding a student’s subscale performance level is to estimate the cut score (or performance standard) at the subscale level. Most assessments utilize a performance standard at the overall test level to assign students to proficiency categories (e.g. Below Standard, Met Standard, etc.). However, in order to evaluate student performance at the subscale level, which is required to provide accurate diagnostic feedback, it is necessary to have a cut score for that subset of items. There are various methods for estimating subscale cut scores. The Learning Locator service can utilize one of several methods for subscale cut score estimation.

In reviewing the methods used in various state testing programs, (California, Oregon, Nevada, Utah, Arizona, New Mexico, Colorado, Kansas, Mississippi, Illinois, Michigan, Tennessee, North Carolina, Maryland, Massachusetts, New York, New Jersey, Wisconsin, Virginia and Vermont) two basic types of estimation procedures were most widely reported depending on the underlying score metric of the assessment. If the subscale scores are IRT-based (i.e. on the theta scale), the theta value, or ability estimate, associated with the overall test cut score is typically applied to each subscale score to classify students. If the score is not on the IRT theta scale, the subscale cut score is frequently calculated by taking all students who achieved just above a “proficient” score on the overall test and finding the average performance of
those students on the subscale of interest. Once subscale performance and subscale cut scores have been estimated, student subscale performance can be evaluated. The Learning Locator can implement either of these two methods.

Choosing a proper method for establishing cut scores for the subscales depends on several practical and technical factors. The practical factors include (1) state policy, (2) the IRT model, (3) the subscale score metric and the need to compute a subscale cut score, and (4) the administration mode of a test. The technical factors include (1) accuracy, (2) score reliability, and (3) classification accuracy. Because each assessment is unique, Pearson works with each client to determine the best method for their specific needs.

Implementing an appropriate method to estimate subscale cut scores ensures that each student is provided with materials that meet their learning needs. Both the subscale estimation and cut score estimation procedures are essential components of step 3, the generation of a student profile.

**Step 3: Generation of a Student Learning Profile**

After estimating the student subscale score and comparing it to the estimated subscale cut scores for each objective, Perspective generates a student learning profile for each student. Perspective utilizes subscale score performance to generate a student learning profile. The student profile indicates student performance relative to the Proficient cut score using a three, four, or five point scale (depending on client preference). That student learning profile is then associated with an encoded record identification number (Learning Locator). That Learning Locator is then printed on the Individual Student Report delivered to each student following the state
assessment. Students receive a Learning Locator for each subject in which they tested. On the Perspective website, teachers, families, and students can enter the code in the Learning Locator field and access resources specific to that student's needs.

Subscale scores are frequently used to create student learning profiles. A profile is essentially a pattern of test scores on different tests or on test objectives within the same test. Mehrens & Lehmann (1973) define a profile as “a graphic representation of the results on several tests, for either an individual or a group, when the results have been expressed in some uniform or comparable terms (standard scores, percentile ranks, grade equivalents, etc.) The profile method of presentation permits identification of areas of strength or weakness.” A profile is generated when a collection of test scores or objective scores are placed on “a graph or chart, side by side, using the same scale for all of them so that comparisons can be visualized.” (Hills, 1993).

Test profiles have been generated and analyzed for over 50 years. Cronbach and Gleser (1953) are credited with the first research devoted to the psychometric and statistical characteristics of profiles and though little research has been conducted since, Brennan (2005) has done some work in this area.

Profiles are useful not only for summarizing and interpreting student performance, but for pinpointing areas in which students need additional attention. For instance, a profile might indicate that a student is able to solve algebraic equations with one unknown but struggles with solving equations with two unknowns. That information can then be used to create a plan of
remediation to narrow the gap between the student’s current achievement level and his or her target level. A profile is essentially a blueprint of student performance on the various subscales.

One of the reasons that profiles are effective is their structure. The profile itself mimics the structure of the assessment and of the curriculum being measured on the assessment. For example, if the mathematics curriculum contains an objective on “Numbers, Operations, and Quantitative Reasoning,” that objective will be presented as a subscale component of the student’s learning profile. Profiles are handy summaries of student strengths and weaknesses generated from the assessment; they help align the assessment with instruction and learning.

To help understand how learning profiles work, an analogy to health screening and medical evaluation procedures might be useful. Consider two children who suffer from excessive fatigue. They have both been sleeping more than normal and report having no energy to engage in the activities they normally enjoy. Upon visiting their pediatrician, the doctor conducts his normal physical examination by looking into the eyes, ears, nose and throat and palpitating the lymph nodes to see if they are swollen. The physician takes blood pressure and temperature readings. Unable to detect any health problems from the physical examination, the doctor orders blood tests to see if they will reveal any underlying cause for the symptoms. The blood work order includes measurement of the complete blood count, blood sugar level, kidney and liver function, blood cholesterol levels, and thyroid function.
Child A’s results reveal a high white blood cell count, indicating that he is fighting off either a viral or bacterial infection. Using the blood test results, the doctor then recommends a course of action. In the event that the infection is bacterial, he prescribes a round of antibiotics. He also prescribes drinking lots of fluids, increased consumption of vitamin C, and bed rest so that the child’s body can stave off the infection.

On the other hand, child B’s results indicate normal blood cell counts, but a low thyroid level. This child does not have an infection, but rather a genetic condition in which an endocrine gland is not functioning properly, that is leading to his fatigued state. Subsequently, the doctor prescribes thyroid medication to restore the child’s levels to within a normal range. After a few days of treatment, the child’s energy level returns. In this analogy, the blood test results are considered to be a “profile” of the child’s health. Just like the health profile is used to diagnose a medical condition and direct the patient to different treatments, the student test results in the form of a learning profile are used to route the student to different areas of attention. As is the case in this example, two profiles that are similar in nature can lead to different remedies. Two similar student profiles can lead to two different sets of learning resources just as two similar health profiles can lead to different medical treatments.

**Step 4: Link to Learning Resources**

The Learning Locator links student learning profiles (in the form of a record identification number) for each student’s performance to a repository of student resources and interactive student lessons targeted specifically for that
assessment profile. Learning resources are chosen by Content Specialists who have been carefully selected based on their teaching experience, training in the Perspective system, and training on the curriculum standards of the client. A variety of supplemental learning resources is reviewed in order to provide learning activities for all of the assessed standards on a test. The learning resources are selected by Content Specialists that:

- Are directly linked to the assessed standard and the student's specific achievement level
- Contain quality content
- Are appealing to students
- Are presented to students in a variety of formats (games, worksheets, guided lessons, quizzes)

Resources that meet the above criteria are included in the grade level pool of resources and are linked to the assessed benchmark through content alignment. Content alignment activities are conducted by Content Specialists. The outcome of the alignment activity is a set of catalogued resources that are specifically correlated to the state assessment and to the student's achievement level. This pool of resources is then reviewed and narrowed in order to help students focus on resources that will provide the most effective practice in specific standards and at a particular performance level. Learning resources are correlated directly to a student's assessment profile through the Learning Locator. Because the resources are assigned based on a
student's specific assessment profile, some student profiles will include resources that are below or above the student's enrolled grade level.
Summary: Implementation of the Learning Locator

The Learning Locator is the mechanism that connects students with appropriate learning materials based on their assessment performance. A student profile is generated based on test performance and that profile is then associated with a Learning Locator. Learning resources are catalogued by Content Specialists and associated with performance profiles. The Learning Locator code is then used to directly match a student with the appropriate materials. The coupling of sophisticated statistical techniques and content expert judgment results in personalized learning resources linked to the most appropriate resources for each student’s needs.

The Perspective system resides within the context of a specific statewide test. Within the Perspective system, the mechanism called Learning Locator drives the integration of the remaining three components: customized reports, online portal, and learning resources. Together, the parts of the Perspective System integrate the elements of assessment, instruction, and student learning. Through the coordinated integration of these elements, Perspective leverages assessment information to give teachers and families ready access to the tools developed to increase student learning.
Perspective Example
Using Perspective’s Learning Locator is intuitive for students, families, and teachers. This section walks through an illustrated example of how to use the service. After a student takes the assessment, an individualized Student Report (ISR) is generated. A generic ISR looks like the following:

![Image of Sample Student Report](image.png)

Figure 2: Example of an Individualized Student Report
For each subject on the score report (e.g. Reading and Mathematics), all basic client requested score information (scaled scores, proficiency levels, performance relative to campus, district, state peers, etc.) is displayed. In addition, a Learning Locator is printed for each subject based on the student’s performance as discussed previously. For example, in the figure above, the student earned a scaled score of 260 on Mathematics which falls into the “Basic” performance category. Based on that student’s specific performance on the various assessment objectives, she was placed in the specific group to which Learning Locator code M103562 was assigned.

Once the student and parent receive the Individualized Score Report (ISR), they can access the Pearson Perspective website and enter the Learning Locator code to access their personalized learning resources. Below is a snapshot of the Perspective webpage and illustration of how to enter the Learning Locator code from the ISR.
Figure 3: Perspective Landing Page Students enter their Learning Locator on the landing page’s data entry field.

The Learning Locator from the ISR is simply entered into the orange box on the left-hand side of the screen. Then, by clicking on the “Show Materials” button, the Learning Locator returns supplemental learning resources based on the performance level information associated with the Learning Locator code. An example of those resources for grade 3 Mathematics is shown below.
Figure 4: Personalized Resources Display of resources appropriate to that student’s test performance profile.
Next, students, families, or educators can simply click on any of the hyperlinks (in blue) to open a variety of interactive lessons, PDFs, learning games, videos, quizzes, or other materials targeted specifically to the student’s performance level. For example, clicking on “Missing Numbers” directs the user to the following activity:

Figure 5: Example of “Missing Numbers” Exercise
Conclusion

Learning Locator is an integral part of the Perspective integrated assessment system. Utilizing subscale estimation, subscale cut score estimation techniques, and student learning profile information, the code takes student assessment performance and links it to a collection of learning resources unique to the needs of each student. The strengths of the Learning Locator system stem from its scientifically based approach. Student learning profiles are generating using statistical and psychometric models. In addition, Content Specialists use content alignment methods to map test performance to instructional materials. The Perspective system links assessment and instruction in the context of large scale testing, replicating the time-consuming process used in the classroom. Learning Locator is a vehicle that encourages students to successfully bridge the gap between their current achievement level and their required one.

Efforts are currently underway to evaluate the success of the Perspective system and the Learning Locator product. It is currently deployed in Minnesota and feedback has been positive. Research is ongoing to evaluate whether teachers and students are receiving the information they need from the Perspective system.

Several studies are planned aimed at evaluating the Learning Locator product in terms of the consequences (Messick, 1989) for the student, teacher, and parent. In future years, surveys will be administered to teachers, students, and parents who have used the system to assess the degree to which they feel it has improved their teaching and learning. Students will be asked
to evaluate the degree to which use of the targeted resources has helped them better understand the targeted concepts on the state assessment. Parents will be asked to evaluate the degree to which they were better able to help their child address his or her learning needs as a result of using the system. In addition, teachers will be asked to evaluate the degree to which the Perspective system impacted any classroom teaching practices and to assess the usefulness of each feature of the Perspective system. In addition to the survey data, actual student performance data will be collected and analyzed to determine the impact of the system on student achievement. This ongoing research will be used to continually evaluate the Perspective and Learning Locator products to guide enhancements and improve the utility of the system.
References


Appendix B: Learning Locator Features

The Learning Locator service provides extensive features and functionality. Among those are the following:

1. The Learning Locator code definition is completely customizable. It can be configured to generate codes only within a pre-defined range or can be set to expire at a given date. Additionally, exclusion rules can be defined in the system such that students omitting a certain number of items will not receive a score report.

2. The Learning Locator system can support multiple test administrations for different years.

3. The Learning Locator code length can be reduced by converting the code into hexadecimal codes. Therefore, instead of having a 7 digit record length, it can be shortened, if so desired by the client, while still maintaining accuracy and security. This feature provides greater flexibility in report design and facilitates use by younger students.

4. The Learning Locator system supports different numbers of performance levels. These performance levels are defined by the client and based on the performance levels used in the operational assessment. For example, an assessment can yield performance levels of Below Standard/ Met Standard or levels of Basic/Proficient/Advanced. The number and names of the proficiency levels are tailored to the client needs. These proficiency levels are then applied to each subscale within the assessment.
5. Learning Locator can support multiple objectives; any number of objectives measured within a subject area can be supported by the Learning Locator service.

6. The learning resources are aligned to the academic standards used by the client. The learning resources retrieved are at, above, and below grade level. All resources are linked to the state standards specified by the client and measured on the assessment. In order to meet the needs of all students, some resources will be at the enrolled grade level, while some will be above or below the targeted grade level.
Appendix B: Subscale Score Estimation Methods

The Bock Method
The data from the simulation procedure were used to estimate the examinees’ \( \theta \) values for the whole test. The estimated parameters, \( \hat{\theta} \), were then entered into the equation (a) \( P_y(\hat{\theta}) = c_y + (1 - c_y) \frac{\exp[1.7a_y(\hat{\theta} - b_y)]}{1 + \exp[1.7a_y(\hat{\theta} - b_y)]} \) or (b) \( P_{yk}(\hat{\theta}) = \frac{\exp\left[\sum_{v=1}^{k} a_y(\hat{\theta} - b_{yv})\right]}{\sum_{c=0}^{m} \exp\left[\sum_{v=0}^{c} a_y(\hat{\theta} - b_{yv})\right]} \) to estimate \( P_y(\hat{\theta}) \) and \( P_{yk}(\hat{\theta}) \), where \( i, j, v, k, \) and \( m_i \) represent item, objective, score level index, current computed score level, and total score levels for item \( i \), respectively. The objective scores for objective \( j \), IRT \( T \), were then computed by the equation

\[
IRT_{Tj} = \frac{1}{n_j} \sum_{i=1}^{l_j} \varepsilon_{ij} \hat{\theta},
\]

where \( i \) represents item, \( l_j \) is the number of items in objective \( j \), and \( n_j \) is the maximum possible points in objective \( j \). Note that \( n_j \) equals \( \sum_{i=1}^{l_j} (m_i - 1) \).

For MC items,

\[
\varepsilon_{ij}(\hat{\theta}) = P_y(\hat{\theta});
\]

for CR items,

\[
\varepsilon_{ij}(\hat{\theta}) = \sum_{k=1}^{m_i} (k - 1)P_{yk}(\hat{\theta}).
\]

Bayes estimates of scale scores, \( \hat{\theta} \), were used to compute the IRT objective scores with a normal (0, 1) as the prior distribution for the abilities. Usually when the objective scores are estimated, the item parameters (i.e., the item pool) already exist. Therefore, in this study, the item parameters were assumed known.

The variance of the IRT \( T \) can be expressed in the equation below:

\[
VAR_{IRT_{Tj}} = \frac{\sum_{i=1}^{n_{mc}} P_y(\hat{\theta})[1 - P_y(\hat{\theta})] + \sum_{i=1}^{n_{cr}} \left\{ \sum_{k=1}^{m_i} (k - 1)^2 P_{yk}(\hat{\theta}) \right\} - \sum_{k=1}^{m_i} (k - 1)P_{yk}(\hat{\theta})}{n_j^2},
\]

where \( n_{mc} \) and \( n_{cr} \) are the number of MC and CR items, respectively, and \( n_j \) represents, as defined previously, the maximum possible points in the objective \( j \).
The Yen Method

The following steps were used to estimate Yen’s T (Yen et al., 1997):

1. Estimate IRT item parameters for all selected items
2. Estimate $T$ for the whole test (including all objectives).
3. For each objective, calculate IRT $T_j$ (see equation 1 on page 9), $\hat{T}_j$, where $j$ represents objective $j$.
4. Obtain

$$Q = \sum_{j=1}^J n_j \frac{(\frac{x_j}{n_j} - \hat{T}_j)^2}{T_j(1 - T_j)}.$$  

If $Q > \chi^2(J, 10)$, then Yen $T_j$, $\bar{T}_j = \frac{x_j}{n_j}$, $p_j = \frac{x_j}{n_j}$, (6) and

$$q_j = n_j - x_j,$$  (7)

where $x_j$ is the observed score obtained in objective $j$, $n_j$ is the max points that can be obtained in objective $j$, and $J$ is the number of objectives.

If $Q \leq \chi^2(J, 10)$, then

$$p_j = \hat{T}_jn_j^* + x_j,$$  (8)

and

$$q_j = (1 - \hat{T}_j)n_j^* + n_j - x_j.$$  (9)

The Yen $T_j$, $\bar{T}_j$, is defined as $\bar{T}_j = \frac{p_j}{p_j + q_j} = \frac{\hat{T}_jn_j^* + x_j}{n_j^* + n_j}$, where

$$n_j^* = \frac{\mu(T_j | \theta)[1 - \mu(T_j | \theta)]}{\sigma^2(T_j | \theta)} - 1$$

$$\approx \left\{ \left[ \frac{1}{n_j} \sum_{i=1}^{t_i} \epsilon_i \hat{y}(\hat{\theta}) \right] \left[ 1 - \frac{1}{n_j} \sum_{i=1}^{t_i} \epsilon_i \hat{y}(\hat{\theta}) \right] \right\} \frac{1}{\frac{1}{n_j} \sum_{i=1}^{t_i} \epsilon_i \hat{y}'(\hat{\theta})} - 1.$$  (10)

For MC items,

$$\epsilon_i \hat{y}'(\hat{\theta}) = \frac{1.7a_i [1 - P_i(\hat{\theta})] [P_i(\hat{\theta}) - c_{ij}]}{(1 - c_{ij})}$$  (11)

and
where \( n_{MC} \) is the total number of MC items in the test. For CR items,

\[
e_{ij}^\prime(\hat{\theta}) = a_{ij} \sum_{k=1}^{m} e_{ij}(\hat{\theta}) \left[ (k-1) - e_{ij}(\hat{\theta}) \right]
\]

\[
= a_{ij} \sum_{k=1}^{m} \left( (k-1)^2 P_{ijk}(\hat{\theta}) - \left[ e_{ij}(\hat{\theta}) \right]^2 \right)
\]

and

\[
I(\theta, \hat{\theta}) = \sum_{i=1}^{n_{MC}} \left[ e_{ij}^\prime(\hat{\theta}) \right]^2
\]

\[
= \sum_{i=1}^{n_{MC}} \frac{1}{1 - c_{ij}} \left[ 1 - P_{ij}(\hat{\theta}) \right] \left[ P_{ij}(\hat{\theta}) - c_{ij} \right]^2
\]

where \( n_{MC} \) is the total number of CR items in the test. The definition of \( e_{ij}(\hat{\theta}) \) is in equations 2 and 3.

The variance of Yen T can be expressed in the equation below:

\[
\text{Var}_{Yen T} = \frac{p_j q_j}{(p_j + q_j)^2 (p_j + q_j + 1)} ,
\]

where \( p_j \) and \( q_j \) are defined in equations 6 through 9.

**The Wainer Method**

In vector notation, for the multivariate situation involving several objective scores collected in the vector \( \mathbf{x} \),

\[
\text{REG T} = \mathbf{x} + \beta (\mathbf{x} - \mathbf{x}) ,
\]

\( \mathbf{x} \) is the mean vector that contains the means of each objective involved. \( \beta \) is a matrix that is the multivariate analog for the estimated reliability for each objective. All that is needed to calculate REG T are estimates of \( \beta \). The equation for calculating \( \beta \) is
where $\Sigma_{obs}$ is the observed variance-covariance matrix of the objective scores and $\Sigma_{true}$ is the variance-covariance matrix of the true objective scores.

Because errors are uncorrelated with true scores, it is easy to see that

$$
\Sigma_{true} = \Sigma_{obs},
$$

where $\tau_{xy}$ and $\tau_{x'y'}$ are the off-diagonal elements of $\Sigma_{true}$ and $\Sigma_{obs}$, the population variance-covariance matrices of the true objective scores and observed objective scores. It is in the diagonal elements of $\Sigma_{obs}$ and $\Sigma_{true}$ that the difference arises. However, if the diagonal elements of $\Sigma_{obs}$ are multiplied by the reliability, $\sigma_r^2 / \sigma_x^2$, of the subscale in question, the results are the diagonal elements of $\Sigma_{true}$. It is customary to estimate reliability with Cronbach’s coefficient $\alpha$ (Wainer et al., 2000).

The score variance of the estimates for the $v$th objective is the $v$th diagonal element of the matrix,

$$
C = \Sigma_{true}^{-1} \Sigma_{true}
$$

The IRT TCC Method

Procedures

1. Find the initial value of $\theta$. The initial value of $\theta$ is

$$
\theta^0_r = \ln \left[ \frac{r}{(M - r)} \right],
$$

where $\theta^0_r$ is the initial $\theta$ value for raw score $r$, and $M$ is the possible maximum score for the subscale of interest.

2. Use the Newton-Raphson method to find the $\theta$ value corresponding raw score $r$.

$$
\theta^+_r = \theta^-_r - \frac{func(\theta^-_r)}{func'(\theta^-_r)},
$$

where

$$
func(\theta^-_r) = r - \sum_{j \in X} P_{rjk} \left( \theta^-_r ; \beta^-_{jk} \right),
$$

and

$$
func'(\theta^-_r) = -\sum_{j \in X} P'_{rjk} \left( \theta^-_r ; \beta^-_{jk} \right).
$$
In Equations (21) and (22), the summation $j:X$ is over items on test $X$; $\beta_{jk}$ represents the item parameter vector for item $j$ and score category $k$; $P_{rk}(\theta_r; \beta_{jk})$ is the probability that examinees of a given ability, $\theta_r$, will answer a particular item $j$ correctly for MC items, or will obtain a particular score $k$ for CR items.

3. The procedures requires some iterations and a convergence criterion for successive values of $\theta_r$ such as

$$|\theta_r^+ - \theta_r^-| < .01.$$  \hspace{1cm} (23)

4. When the convergence criterion is reached, then the estimated ability, $\hat{\theta}_r$, is the last value of $\theta_r$, namely

$$\hat{\theta}_r = \theta_r^+.$$  \hspace{1cm} (24)

with variance

$$Var(\theta_r) = 1/\text{Inf}(\theta_r),$$  \hspace{1cm} (25)

where $\text{Inf}(\theta_r)$ is the test information function.

**The IRT EAP Method**

The equation for computing the expected a posteriori estimation method (EAP) is

$$\hat{\theta}_{\text{EAP}} = \frac{\sum_{k=1}^{q} \theta_k L(\theta_k) \phi(\theta_k)}{\sum_{k=1}^{q} L(\theta_k) \phi(\theta_k)}$$  \hspace{1cm} (26)

and the equation for computing the variance of the EAP is

$$S_{\theta_{\text{EAP}}} = \left[ \frac{\sum_{k=1}^{q} (\theta_k - \hat{\theta}_{\text{EAP}})^2 L(\theta_k) \phi(\theta_k)}{\sum_{k=1}^{q} L(\theta_k) \phi(\theta_k)} \right]^{1/2}$$  \hspace{1cm} (27)

where $\theta_k$ is one of the $q$ equidistant quadrature points comprised between $\theta_{\text{min}}$ and $\theta_{\text{max}}$, $L(\theta_k)$ is the likelihood of the responses pattern, and $\phi(\theta_k)$ is the weight associated to each of the quadrature point according to a standardized normal distribution.
**Objective Score Estimation Comparison**

In order to compare the performance of these various estimation procedures, Shin (2005) conducted an empirical investigation. An adjusted version of Bock et al.’s item response theory (IRT) approach (Bock et al., 1997), Yen’s OPI approach (Yen, 1997), Wainer et al.’s regressed score approach (Wainer et al., 2000), Shin’s MCMC regressed score approach (Shin et al., 2005), and the proportion-correct score approach were compared.

Shin’s (2005) study indicated that the methods utilizing collateral test information to estimate subscale scores increased the reliability and precision (in terms of the width of the confidence interval around scores and statistical bias indices). Specifically, the Wainer and Shin methods yielded subscale scores with the highest reliability, while the Yen method was the most precise. The factors that affected reliability were the number of items measuring each testing objective, the correlation between objectives, and the ratio of constructed response items to multiple choice items. The higher the value of these variables, the more reliable the estimated subscale scores were. Within the methods studied, the Wainer and Shin methods yielded the least biased subscale score estimates. The estimates were impacted by the number of items per objective and the correlation between objective scores. The more items per objective or the higher the correlation between subscale scores, the smaller the measures of bias.

As the only empirical comparison of the various subscale score estimation procedures to date, Shin’s 2005 study revealed the relative strengths and weaknesses of each method and the conditions under which each method operated optimally. Because every testing program has unique needs, flexibility in choosing a subscale score estimation procedure and understanding the potential impact associated with each method is critical. This research formed the basis for construction of the Learning Locator algorithm.

For information on Pearson’s complement of assessment and research services, please visit http://www.pearsonedmeasurement.com.

Copyright © 2008 Pearson Education, Inc. or its affiliate(s). All rights reserved.