VIII. Tasks for Diagnostic Assessments to Support the GCA

DRC is proposing a fully customized solution for the Diagnostic Assessment Tool, which we believe will better serve the long-term, rather than the short-term, interests of PDE. In lieu of having a pool of generic items that have been developed for generic purposes, based on generic standards, DRC proposes to develop items specifically for Pennsylvania. We will draw on our extensive experience to align these items to the Commonwealth's content standards and to ensure that they meet the needs of Pennsylvania's educational and assessment systems. Therefore, these items will be administered to students exclusively in Pennsylvania. It may not be the simplest or easiest route to fulfilling the requirements in the RFP; however, we believe that it is the best solution to develop a diagnostic tool of the highest quality.

It is in this spirit that DRC responds to the technical requirements of the RFP and the future implementation of the Diagnostic Assessment Tool. We look forward to the opportunity to work collaboratively with PDE to design this assessment program.

VIII.A. Item Development

DRC and WestEd will provide an extensive multiple-choice item bank with items (including mostly commissioned passages, graphics, and scenarios as appropriate) for all assessments to support the Diagnostic Assessment Tool. The staffs of DRC and WestEd will develop all the items for Algebra I, Algebra II, Geometry, Biology, Chemistry, Literature, English Composition, U.S. History, World History, and Civics & Government aligned to the Pennsylvania Assessment Anchors. We will provide multiple-choice items written to Webb's depth-of-knowledge levels 1, 2, and 3 that match the content and are aligned to the Model Curriculum.

As with the item development plan for the GCA exams, the item development plan for the Diagnostic Assessment Tool will follow the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1999) since items are developed to reflect the range of cognitive ability inherent in the standards, resulting in reliable and instructionally valid tests. In addition, our Diagnostic Assessment Tool item development plan also adheres to the Principles of Universal Design and reflects that we clearly understand how items and tests must lend themselves to accessibility by diverse groups of students and function appropriately across a broad range of test administration accommodations.

The item development processes employed for the items of the Diagnostic Assessment Tool will mirror the GCA development process, with the exception of the committee reviews (item and bias reviews), which are not required by the

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

RFP. To ensure item quality, alignment to the Pennsylvania Assessment Anchors, and alignment to the GGA and the Model Curriculum, we propose the following item development process:

Item Development Process

- Meet with PDE: item development planning meeting
- Select and train item writers
- Development of items, passages, prompts, and scenarios including graphics
- Internal review process will include, but not be limited to:
  - Content alignment.
  - Grade-level appropriateness.
  - Rigor.
  - Depth-of-knowledge.
  - Universal Design.
  - Technical design.
  - Prerequisite skills.
  - Bias, sensitivity, and fairness.
- Select items for field testing; submit selections to PDE for approval
- Post-field test data review (optional)

The costs for the optional post-field test (item/data) review are provided in our Cost Submittal. For more detailed information about item development for the Diagnostic Assessment Tool items, please see *Subheading VII.B* in the GCA work plan of this proposal, as well as *Appendix 2*.

VIII.A.1. Item Bank

Items, including all associated stimulus prompts, scenarios, graphics, passages, and item characteristics will be entered into IDEAS, DRC's item-banking system. Please see *Subheading VII.B* for a detailed description of IDEAS.

PDE staff will have secure, remote, 24/7 access to IDEAS in order to search and view items and passages, as well as print item cards. The system can also be used to facilitate discussion between DRC and PDE regarding items and passages. DRC staff members will be able to attach flags to items and passages; these flags will be readily viewable and searchable by PDE staff and will indicate the need for PDE review and response. PDE staff will have the capability to insert questions or proposed revisions as comments and associate them with individual

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

items and/or passages. When items or passages need revision, PDE will be able to print the specific item cards, make edits, and send to the appropriate DRC team members. This method will provide for a collaborative and efficient item/passage review and revision process.

Item banks will be developed for the four content areas: Language Arts, Mathematics, Science, and Social Studies. These items, written across the grade spans (starting in grade 6), will constitute the item bank for each of the four content areas. Because English Composition and Physics will not be assessed in the Diagnostic Assessment Tool, there will be no items in the item bank for these areas. A diagnostic supplement for English Composition will be provided along with the other constructed-response diagnostic supplements. See *Subheading VIII.A.4* for the constructed-response supplement plans. Physics, though part of the Model Curriculum, does not have a GCA component. Thus, Physics items are not needed for the Diagnostic Assessment Tool.

The operational item bank will form the item pool that will be used to select items for each Diagnostic Assessment Tool exam. Note that the items for the Diagnostic Assessment Tool will be written to assess the knowledge, skills, and abilities as outlined in the Assessment Anchors and as taught in the Model Curriculum.

The item counts identified later in this section reflect those that are needed to get the Diagnostic Assessment Tool ready for the scheduled operational release per negotiation discussions with PDE. Should PDE desire more items and an updated item bank (i.e., to annually update the Diagnostic Assessment Tool bank with new items), DRC will work with PDE to add items through an embedded field testing design. This would require additional item development and related costs, which are not included in the current proposal.

With an item bank for each content area, DRC's conceptual design of the Diagnostic Assessment Tool is illustrated in the figure below. The Diagnostic Assessment Tool is a system with four components, one for each content area. Each component (content area) has an exam for each grade or course. The pool of items will be used to create exams administered in a stage-adaptive testing routine. See *Subheading VIII.B.1* for details on DRC and CAL's proposed stage-adaptive testing system for the Diagnostic Assessment Tool. The grades and courses applicable to each content area are listed below. For example, for the mathematics component of the Diagnostic Assessment Tool, students can start taking the exams at grade 6. Results at the lower grades are used to diagnose a student's strengths and needs, as well as to estimate the likelihood of success on the GCA exams (i.e., Algebra I, Algebra II, and Geometry). Students who are currently enrolled in GCA courses can also take the appropriate Diagnostic Assessment Tool exams. Their results will also be used to diagnose their strengths and needs and to predict performance on the GCA exams.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

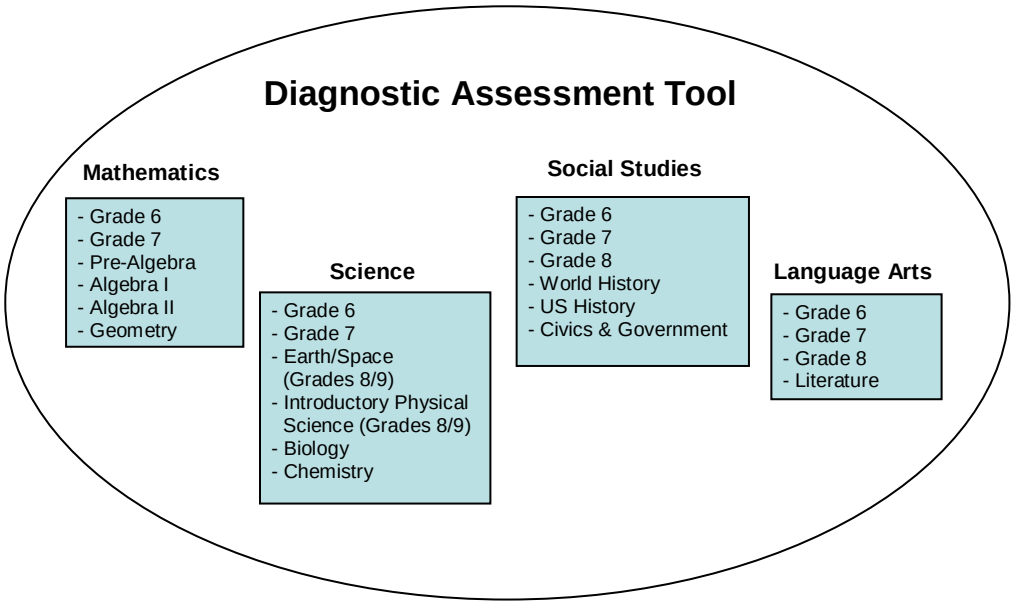


Illustration of the Diagnostic Assessment Tool System Consisting of Four Components (Content Areas)

VIII.A.2. Item Chart by Content Area and/or Grade Level

These charts indicate the number of items that will be field tested by content area and/or grade level, as appropriate.

Mathematics

Course/Grade	Fall 2009
	Number of MC Items
6th	260
7th	260
Pre-Algebra	260
Algebra I	260
Algebra II	260
Geometry	260

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Science

Course/Grade	Fall 2010 Number of MC Items
6th	260
7th	260
Earth/Space Grade 8/9	260
Introductory Physical Science Grade 8/9	260
Biology	260
Chemistry	260

Social Studies

Course/Grade	Fall 2010 Number of MC Items
6th	260
7th	260
8th	260
World History	260
US History	260
Civics & Government	260

Language Arts

Course/Grade	Fall 2010 Number of MC Items
6th	430
7th	430
8th	430
Literature	430
English Composition	430

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.A.3. Item Alignment Chart

As part of the item development process, DRC will provide a chart to identify depth of knowledge, item alignment of Pennsylvania's academic standards, and other information as required by the RFP. Information will be provided for grades 6 through 12 in reading, mathematics, science, and social studies. A separate chart for grades 9 through 12 will be provided for each GCA except English Composition. PDE, DRC, and WestEd will mutually determine a date for when this chart is to be provided.

VIII.A.4. Constructed-Response Supplement

DRC will be pleased to provide a constructed-response supplement to the Diagnostic Assessment Tool. It will include sets of constructed-response items, rubrics, scoring guidelines, and professional development to meet all requirements in Appendix B of the RFP. The items and rubrics will be designed to identify a student's strengths and weaknesses and provide information to guide instruction. In addition, DRC will develop and make available to teachers electronic versions of sample prompts and rubrics designed to enhance student's English Composition performance.

Teachers will be able to score these student responses, but the scores will not be collected or reported. DRC will provide the supplement in PDF format. DRC recommends one constructed response for each Pennsylvania Assessment Anchor. Therefore we plan on providing in the constructed-response samples in the following number of items:

Mathematics: Fall 2009

Content	50/50 4Pt CRs	50/50 2Pt CRs	65/35 4Pt CRs	65/35 3Pt CRs	80/20 4Pt CRs
6th	21	21	21	21	42
7th	21	21	21	21	42
Pre Algebra	21	21	21	21	42
Algebra I	21	21	21	21	42
Algebra II	21	21	21	21	42
Geometry	21	21	21	21	42

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Science: Fall 2010

Content	50/50 4Pt CRs	50/50 2Pt CRs	65/35 4Pt CRs	65/35 3Pt CRs	80/20 4Pt CRs
6th	18	18	18	18	36
7th	18	18	18	18	36
Earth/Space Grade 8/9	18	18	18	18	36
Introductory Physical Science Grade 8/9	18	18	18	18	36
Biology	18	18	18	18	36
Chemistry	18	18	18	18	36

Social Studies: Fall 2010

Content	50/50 4Pt CRs	50/50 2Pt CRs	65/35 4Pt CRs	65/35 3Pt CRs	80/20 4Pt CRs
6th	18	18	18	18	36
7th	18	18	18	18	36
8th	18	18	18	18	36
World History	18	18	18	18	36
US History	18	18	18	18	36
Civics & Government	18	18	18	18	36

Language Arts: Fall 2010

Content	50/50 4Pt CR's	50/50 2Pt CR's	65/35 4Pt CR's	65/35 3Pt CR's	80/20 4Pt CR's	Prompts
6th	10	10	10	10	20	12
7th	10	10	10	10	20	12
8th	10	10	10	10	20	12
English Composition	0	0	0	0	0	12
Literature	10	10	10	10	20	0

DRC has provided a PSSA sampler in *Appendix 3* as an example of our previously developed high quality item sampler with constructed response and prompts with rubrics.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.A.5. Statistical and Psychometric Analysis

Statistical and psychometric analyses will be completed for the Diagnostic Assessment Tool in the same manner as the GCA assessment. Please see *Subheading VII.B.9* for a detailed description of analyses to be performed.

VIII.A.6. Meeting the Needs of the Diagnostic Assessment Tool

As stated earlier, DRC and CAL are proposing a customized solution for the Diagnostic Assessment Tool. Items will be developed specifically for this component, based on Pennsylvania Assessment Anchors and aligned to the GCAs and the Model Curriculum. The CAL system will also be configured specifically to meet the needs of Pennsylvania.

VIII.A.7. System for Providing Feedback

As part of the online test system, CAL will include a link within the Management Tools that will allow Pennsylvania educators and support staff to provide general feedback on any aspect or component of the Diagnostic Assessment Tool. The "Feedback" tag will link an authorized user to an interface whereby specific comments from system users can be entered via a text box. Certain fields regarding the nature of the feedback will be pre-defined, as well as an "other" field that can be customized by the user. All feedback will be reviewed by CAL staff in an appropriate timeframe (Priority tags such as Low, Medium, and High will be stamped to each comment by individual users) and will be researched as needed. In instances where action is recommended arising from feedback from the field, the action will be subjected to CAL's internal quality assurance process. Logs and records of all feedback will be available for PDE review on-demand.

VIII.B. ASSESSMENT CONSTRUCTION

VIII.B.1. Testing Algorithms

With CAL's state-of-the-art online testing system, students will have an easy-to-use system to take the Diagnostic Assessment tests. The system will provide immediate diagnostic feedback to students and teachers. Each student's strengths and needs will be diagnosed and immediately available through CAL's diagnostic reporting mechanism. The reports will also include a predictive component that will estimate the likelihood that a student will score in the proficient category on the GCA exam. See *Subheading VIII.J* for details on CAL's diagnostic reporting capabilities.

DRC and CAL are proposing a stage-adaptive testing approach for the Diagnostic Assessment Tool, with items administered in three stages. The figure below provides an illustrative example of how this might work. Note that the item numbers in the figure are for illustrative purposes only, assuming a total test length of 50 items.

In Stage 1, all students in a grade/course will take a common set of 15 items. Based on their performance on these common items, the Diagnostic Assessment

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Tool will compute an ability level (i.e., a theta score based on the Rasch item parameters, which can be transformed into a scale score) for each student. Based on that score, the students will be given one of three item sets in Stage 2. Each item set will differ in their overall mean item difficulty. Item Set 1 in Stage 2 will contain easier items for students who did not do well on the items from Stage 1. Item Set 2 in Stage 2 will contain harder items for students who did moderately well on in Stage 1. Item Set 3 will contain the hardest items for students who performed very well in Stage 1. Similarly, based on their combined performance on the items from Stages 1 and 2, students will be further routed to the appropriate item set in Stage 3.

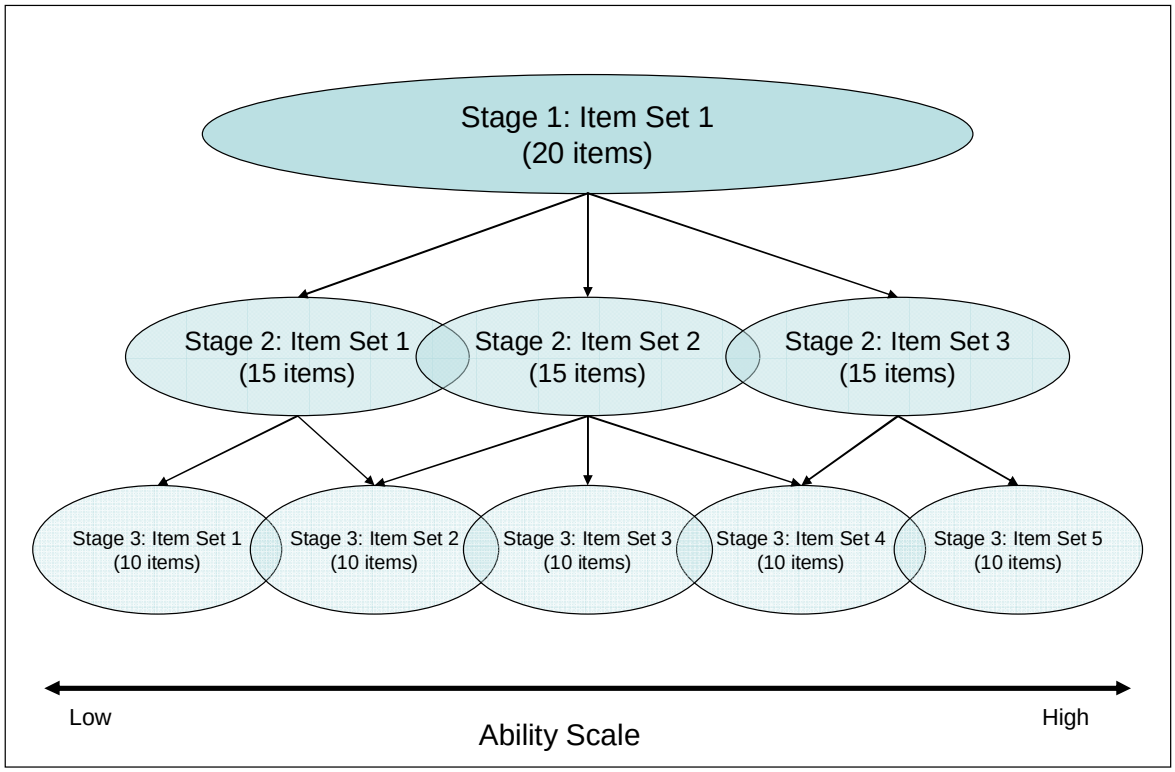


Illustration of the Stage-Adaptive Testing System for One Grade or Course for the Diagnostic Assessment Tool

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

An example of how the stage-adaptive approach would work for an individual student is illustrated below, as indicated by the blue-outlined ovals and blue arrows. In the example, the student performs well (i.e., answers most of the items correctly) in Stage 1, thus earning a high ability score. Based on this score, the student is routed to Stage 2: Item Set 3 items. The student then performs moderately well on the Stage 2 items, and based on the overall performance on the Stage 1 and Stage 2 items, a new ability level is computed for the student. Based on this new ability score, the student is routed to Item Set 4 in Stage 3. After completing the items in Stage 3, the student's final ability and scaled score are computed based on the student's performance on the items from all three stages. In addition to the overall scaled score for the exam, scaled scores will be computed for each measurable skill set (i.e., objective). Note that the number of items in each stage in the figure below is for illustrative purposes only. The exact test specifications for each exam of the Diagnostic Assessment Tool will be determined based on a review of the assessment anchors and eligible content.

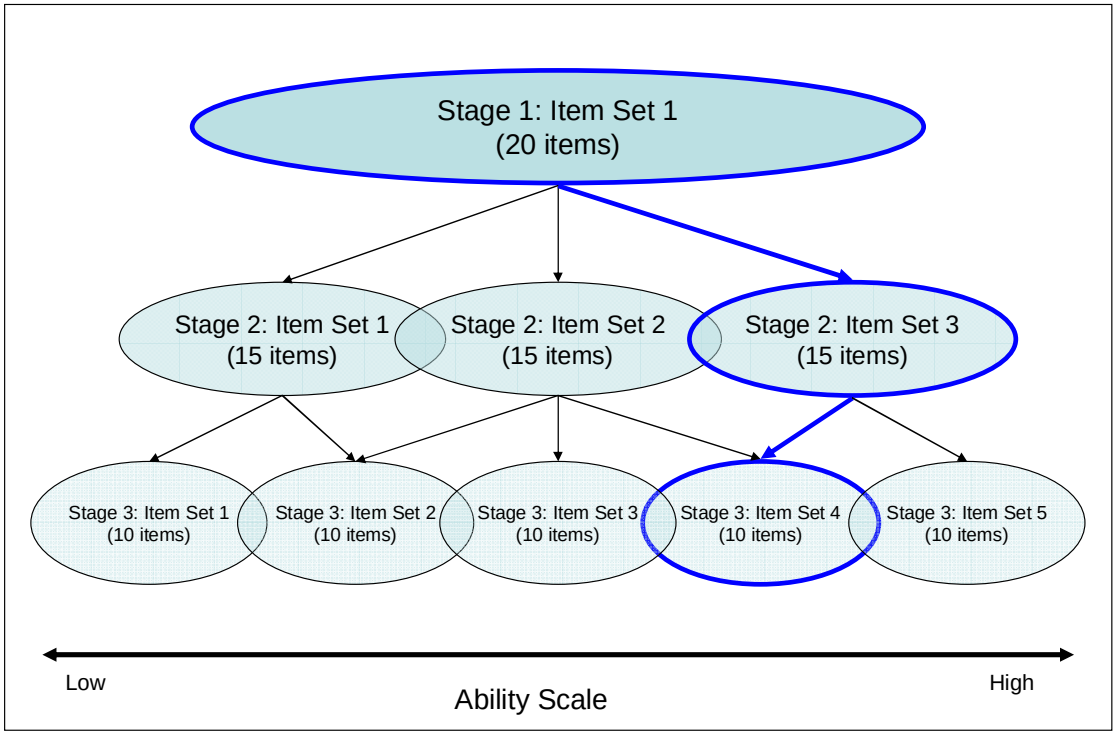


Illustration of the How a Student Progress through the Stage-Adaptive Diagnostic Assessment Tool

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

The stage-adaptive testing algorithm, only used with the operational Diagnostic Assessment Tool (not during field-testing), would use the item selection mechanism illustrated above. A student's abilities, after answering each set of items in each stage, will be computed using a maximum likelihood estimation (MLE) routine. When MLE estimates are not possible (i.e., when a student has answered all the items correctly or incorrectly), a Bayesian estimation routine based on the posterior mean distribution (EAP, or *Expected a Posteriori*) could be computed.

For the operational Diagnostic Assessment Tool exams, the same number of items will be given to each student, making this a *fixed-length* (semi) adaptive exam. In addition to the estimates of ability, standard errors of those estimates will be computed, based on the MLE routine.

For a repeat administration of the exam in the Diagnostic Assessment Tool, a student would take the same set of items in the diagram above, with the following exception. There will be another set of items in Stage 1 (not illustrated). This allows the student to take a different set of items in Stage 1. Items in Stage 2 and 3 will be those illustrated above. But, note that the student would not likely go through the same path along the stage-adaptive routine. Thus, if students have learned the content between test administrations, they will likely traverse down a different path along the stage-adaptive routine (i.e., be administered different items sets). That is, the students will likely see mostly unique items across test administrations.

For this stage-adaptive approach, we will build an item pool of appropriate size for each grade/course in each content area.

VIII.B.2. Online Modular Diagnostic Assessments

As presented above, DRC and CAL are proposing an adaptive test for the online Diagnostic Assessments. Details of the computerized adaptive testing algorithm are provided in *Subheading VIII.B.1*. The adaptive approach is DRC and CAL's preferred plan.

VIII.B.3. Meeting Assessment Specifications

The online assessments for the Diagnostic Assessment Tool will be developed to meet the specifications listed in Appendix B of the RFP. As outlined earlier in our proposal, DRC will develop four item banks, one for each of the four content areas. The items in each content-area item bank will be used to create the exams for each grade or course, beginning in grade 6. See *Subheading VIII.A* for our item development plan for the Diagnostic Assessment Tool. See *Subheading VIII.B.1* for details of the stage-adaptive algorithm that will be used to administer student's exam.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

The test specifications for the GCA require some open-ended (OE) items. However, note that the online Diagnostic Assessment Tool will **not** include any OE items, or prompts/rubrics for the Composition test. See *Subheading VIII.A.4* for DRC's proposed OE and prompt paper-based Diagnostic Tool Supplement.

Ensuring Predictive Validity

An essential feature of the Diagnostic Assessment Tool is its ability to predict future performance on the GCA tests. DRC's solution for Pennsylvania will assist educators in anticipating whether students are on track and have the knowledge and skills needed to achieve at proficient levels on the GCA exams. Simply stated, predictive validity is established by measuring a given characteristic, predicting the outcome of a future event, and waiting to see if the event occurs as predicted. Once predictive validity has been established, the instrument can be used with confidence to accurately assess students on the basis of the expected outcome. To document the predictive utility of the Diagnostic Assessment Tool, comprehensive studies will be performed to ensure the solution is a valid and reliable predictor of GCA performance.

DRC is proposing a comprehensive, ongoing, predictive validity research agenda for the Diagnostic Assessment Tool. The agenda includes the collection of other types of validity evidence (e.g., content and construct evidence), which will be used to help bolster PDE's understanding of any Diagnostic Assessment Tool predictive relationships. Second, the Diagnostic Assessment Tool will be scaled using the Rasch model. This opens up the opportunity to conduct analyses on the Diagnostic Assessment Tool item residuals as well as the development of vertical scales for each Diagnostic Assessment Tool subject area. Both will add value in terms of informing any Diagnostic Assessment Tool longitudinal relationships. In addition, correlation and regression studies are proposed. Further, the regression results will be used to statistically link each Diagnostic Assessment Tool exam to its corresponding GCA. Specific elements of this research agenda are discussed in more detail below.

Background

The term *predictive validity* is used in a variety of assessment contexts. For example, the SAT or ACT assessment is used to predict success in college. In pre-employment testing, one wishes to predict success on the job. In assessment, when one hears the word "prediction," the word likely to follow is "validity."

For the Diagnostic Assessment Tool/GCA situation, the standard conceptualization of predictive validity may not be the most appropriate. Consider the fact that if the system works perfectly, the correlation between the grade six Diagnostic Assessment Tool measures and the GCA measures would be near zero. This condition would occur if the Diagnostic Assessment Tool correctly identifies all students not on track to do well on the GCA so that the teachers, using the Model Curriculum and other tools, could successfully remedy any deficiencies. Then all students should succeed on the GCA, regardless of their earlier Diagnostic Assessment Tool measures. Of course this does not negate the

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

probative value of the proposed line of research. It simply means that other indices may be of more value in considering the broad utility of the Diagnostic Assessment Tool (e.g., the 'diagnostic' values of the Diagnostic Assessment Tool may be significantly more relevant the further away a student is from the GCA criterion).

Predictive validity evidence relates to forecasting success or status at a later point in time. Data for these analyses may be collected in two ways:

1. Present the predictor assessment to students for whom the criterion measure is already available. When the two time points are naturally close together with minimum intervention between (e.g., employment testing), this procedure works well; for a predictor assessment intended for grade six and a criterion from end-of-course high school, it is problematic.
2. Administer the predictor assessment at the intended time and wait until the criterion measure becomes available on the regular schedule. This avoids the validity issues raised for the first method and works well if the time lag is not too great. The Diagnostic Assessment Tool validation would need to accumulate over several years to gather all the required data.

Data for the predictive validity studies can be collected by either of the two methods described above. Method 1 (outlined above) could be used as soon as the tool is in place but would require some out-of-level testing, similar to a vertical scaling study. Method 2 would require one year to collect the initial Data. DRC's plan includes some elements of both.

Methodology

Content and Construct Evidence

The analytic evidence that will be collected as part of the content and construct validity argument can also support a more appropriate approach to predictive validity. Content validity will be established through industry standard processes (e.g., identifying the prerequisites for each academic standard and assessment anchor). Ensuring content comparability between the Diagnostic Assessment Tool and GCA is critical. First, it means that *indicators* for the prerequisite Diagnostic Assessment Tool will be present in the criterion (GCA) assessment. Second, the more similar the content matches between the two assessments, the stronger the resulting statistical linkage.

Rasch Scaling Analysis

Diagnostic Assessment Tool items will be scaled with the Rasch model. Please refer to *Subheading VII.J* for more specific information regarding Rasch scaling. Scaling the Diagnostic Assessment Tool items will allow two supporting studies that fit into DRC's proposed predictive validity research agenda: residual analysis and vertical scaling.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Residual Analysis

Predictive validity evidence for the Diagnostic Assessment Tool can be developed by demonstrating empirically the association between the indicators for the prerequisite and those for the standard per se. The appropriate level of analysis can be at the item or testlet level, as well as the total test. In the context of a Rasch fit analysis, relationships will appear as inter-item associations that are not accounted for by the overall item difficulty but which can be aligned with the prerequisite connections. After extracting the model's sufficient statistics, the predictors and criteria should have positive association in the residuals. These relationships can be investigated and confirmed with a variety of multivariate statistical techniques.

Vertical Scaling

When learning in a subject area is continuous across grades, and instruction is related to increased achievement in the subject area, it is often possible to derive a single score scale that covers the range of content spanning all grade levels. The purpose of vertical scaling is to establish such a score scale score. This concept is represented visually in the figure below. Teachers have long understood that there can be overlapping achievement across grades (e.g., some sixth grade students might spell better than some seventh grade students) as suggested by this figure

A vertical scale should reasonably order both items and students across grades. So, items should be **easier** for students at higher grades and **harder** for students at lower grades. Successful vertical scaling is possible when tests have similar relationships across grades. In other words, the same subject area tests across grade levels are believed to represent a developmental achievement construct. When this is true, each student's status can be located on the continuum in much the same way that each student's height can be located on the appropriate continuum (e.g., centimeters).

Although vertical scales have been successfully developed for a variety of achievement tests, experience has shown that it is not always possible to create educationally meaningful vertical scales. If numerous items are not rank ordered appropriately (e.g., easier for students at higher grades), the construction of a meaningful vertical scale is not possible.

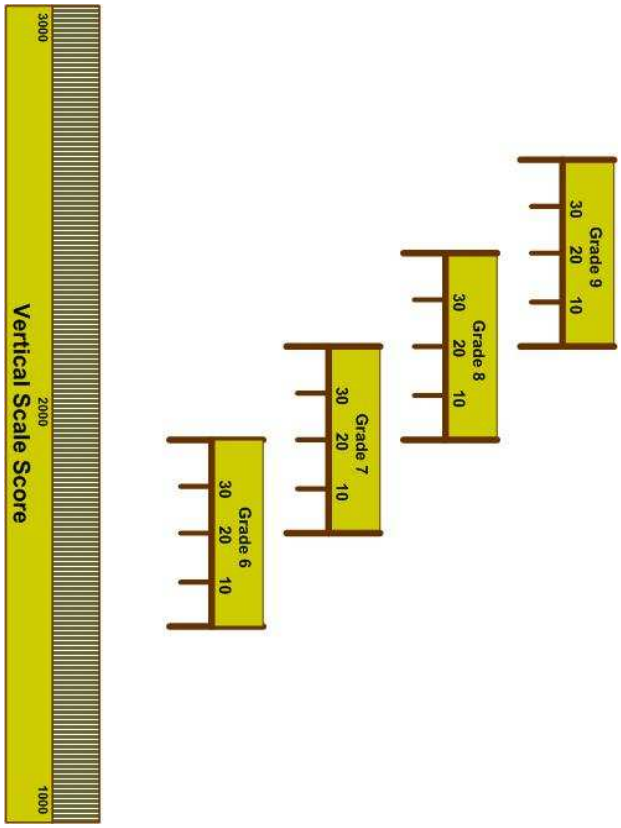
It is a prudent first step to review the scope and sequence of the curricula for each subject area. If grade-eight science is substantively different from grade-nine science, then a vertical scale may not be meaningful. In fact, science can be a problematic subject area when it comes to constructing vertical scales because students often switch among physical, life, and earth sciences across grades.

Other challenges when measuring "out-of-grade" include opportunity to learn issues (e.g., students receive no instruction in some higher grade material) opportunity to forget (e.g., remembering the formula for the area of a rectangle but forgetting the formula for the area of a circle) as well as changes in state (higher-order skills at lower grades can become mechanical processes at higher

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

grades—a quality often attributed to *fractions*). For these and other reasons it is always cautioned that a student who performed well enough on the third grade test to receive a vertical scale score higher than the typical sixth grader does **not** mean the student is ready for sixth grade.

Mapping Scores from Multiple Grades onto a Common Metric.



There are two major procedural components when creating a vertical scale. One involves the data collection design. The second relates to the data analysis methods.

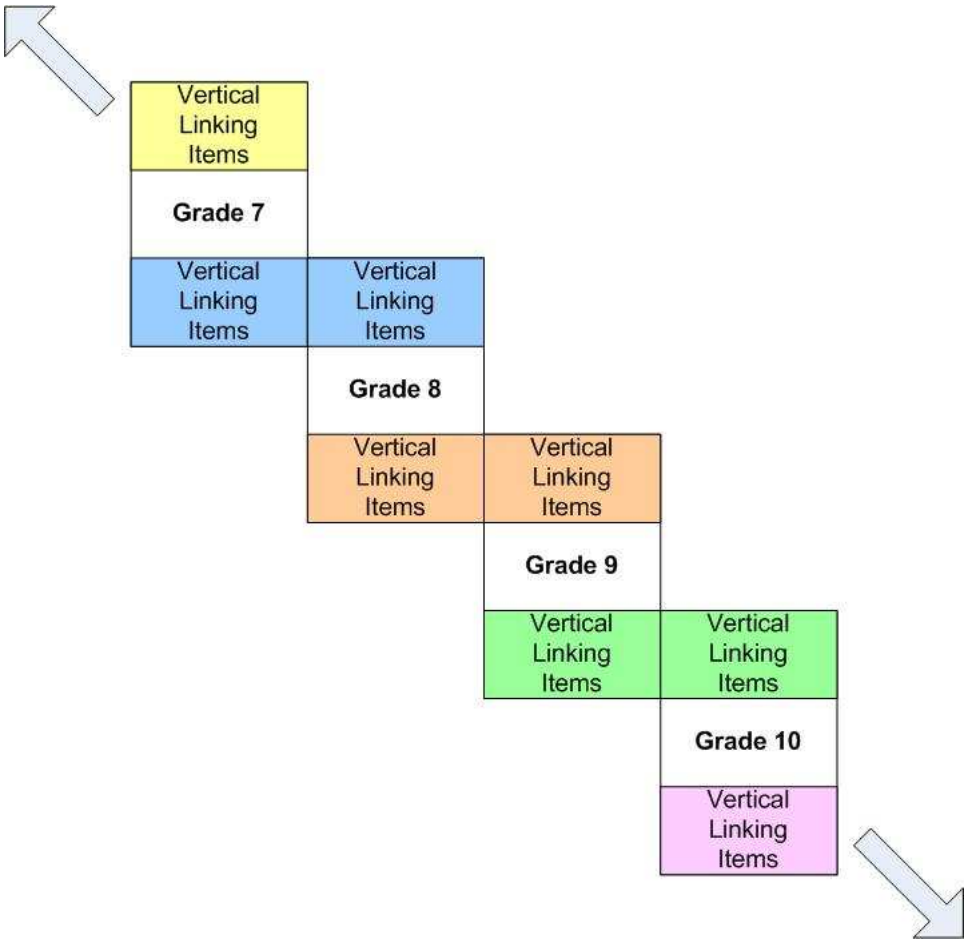
Scaling Design

A common-items nonequivalent groups design is proposed. Each grade and content area Diagnostic Assessment Tool field test will contain embedded vertical linking items. For example, the vertical linking items on each student's grade 7 Diagnostic Assessment Tool Mathematics test will have five (5) Mathematics items from grade 6 and five (5) Mathematics items from grade 8. As illustrated in the next figure, this basic design will be followed until all Diagnostic Assessment Tool grades are connected. (Note that the lowest and highest grades assessed for each subject area will only be able to utilize linking items from the upper or lower grade, respectively). DRC plans to administer the vertical linking items in a matrix format, meaning different sets of students will respond to different sets of

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

vertical linking items (greatly increasing the actual number of vertical linking items).

Illustration of the Vertical Scaling Design.



Scaling Model

Based on DRC's experience and anticipation of good data-to-model fit, the Rasch model will be used to derive the underlying scale score system that will allow the growth of student performance across years to be assessed. This is the same scaling model proposed for the GCAs and is discussed more thoroughly in *Subheading XII.J*. WINSTEPS will be used to analyze the vertical scaling study data.

Proposed Scaling Procedures

The following general steps will be used to create the vertical scale:

1. Separately calibrate the 'on grade' items for each grade and subject area Diagnostic Assessment Tool
2. Use a fixed common item parameter scaling process to place the vertical linking items on the scales created in Step 1

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

3. Evaluate the suitability of vertical linking items (i.e., across grade item-stability checks)
4. Calculate the mean shift needed to put adjoining Diagnostic Assessment Tool levels on same scale
5. Chain all mean shift adjustments together to create a single vertical scale

Special Considerations

An analysis of the suitability of the vertical items is an important prerequisite. Items demonstrating near chance performance (e.g., p-values less than 0.30) or reversals in difficulty across grades will not be not used as vertical linking items.

Typically, a linear transformation is applied to the Rasch ability estimates derived from WINSTEPS to achieve a scale with specific properties. It is beneficial for every raw score to map to a different scale score (although truncation at the extremes—for establishing the lowest and highest possible scale scores—may be needed to keep the range of the vertical scale reasonable). The characteristics of the vertical scale should also be unique from other test scales to avoid confusion on the part of users. Regardless of the final properties of the vertical scale, each subject area Diagnostic Assessment Tool will be linked to the GCA scale. The linking process is described elsewhere in this section.

Sample Tables and Figures

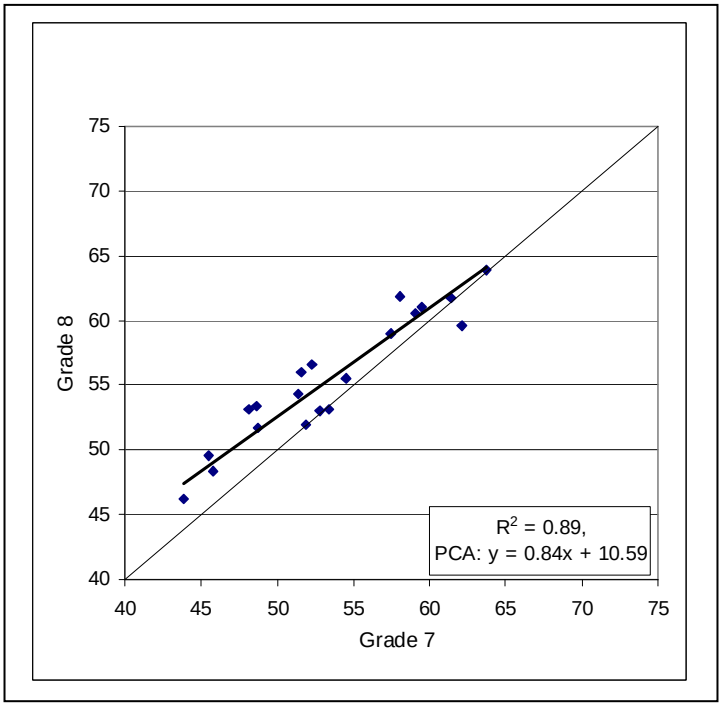
A variety of tables and graphs will be used to describe the results of the scaling study. This information will help describe and visualize the change in year-to-year growth and variability across grades. Several samples tables and figures are provided below.

The next figure demonstrates a procedure used to check across grade item stability. Here the same items were administered at grade 7 and grade 8. The respective *p*-values were transformed using an inverse normal curve function and scaled to have a mean of 50 and a standard deviation of 10. An identity (diagonal) line has been added. Items that fall on the diagonal would have the same difficulty in grade 8 as grade 7. Note that in this example, one item was clearly harder at grade 8 than grade 7. Such items are generally not used in mean shift calculations. A principal components function is plotted representing the 'line of best fit' in the scatter plot. If the change in p-values is essentially uniform, the slope of the PCA line would be about 1, and thus nearly parallel with the identity line.

Similar graphs can be made using the Rasch item difficulty estimates that come from the WINSTEPS program. In fact, WINSTEPS item logit estimates can be used to compute a t-statistic to determine which items do not conform to the model expectations. When sample sizes are large such an inferential test becomes very powerful (meaning trivial differences could lead to statistically significant results). In this case the robust-z procedure can be used to protection against this over-sensitivity.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Sample Across Grade Item Stability Plot



PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Researchers will often rank order item difficulties as in the following table. Upper and lower vertical linking items can be denoted with "#" and "*" signs, respectively. As illustrated in the sample table, these items often turn out to be among the easiest and hardest of the items for any given grade level.

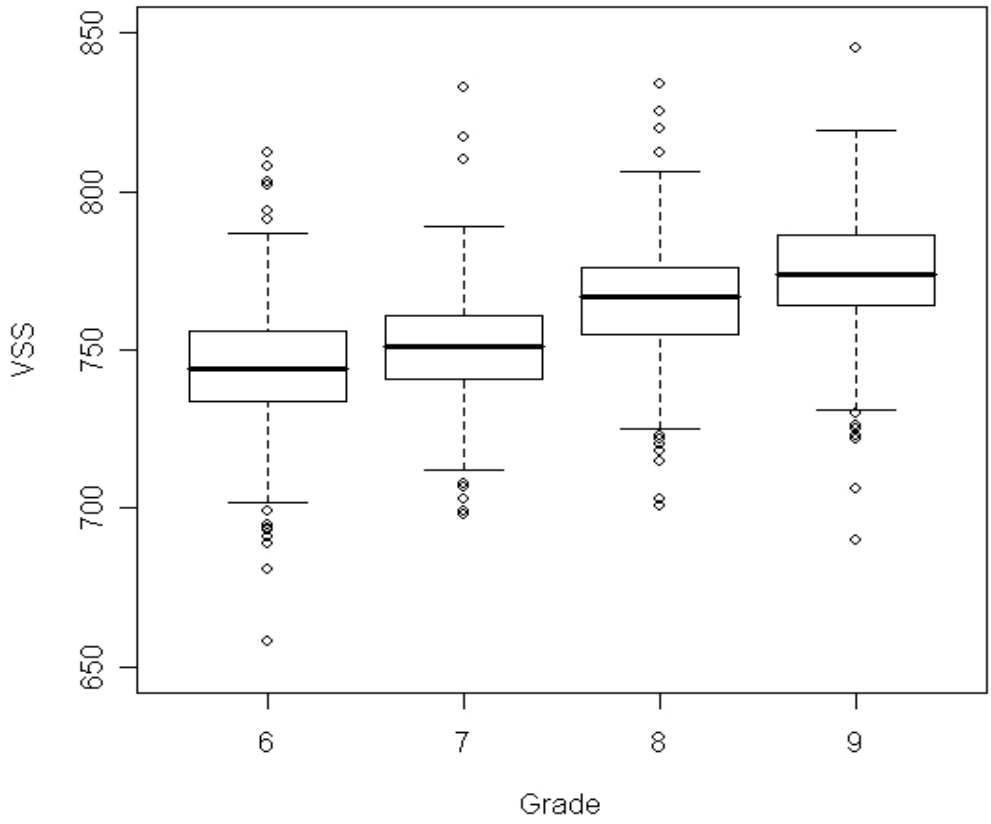
Sample Rank-Ordered Item Difficulties

vlink	entry	measure	error	pval	ptbis				vlink	entry	measure	error	pval	ptbis
*	61	-1.8841	0.0192	0.95	0.26					9	0.0083	0.0077	0.76	0.40
	1	-1.6878	0.0125	0.93	0.29					23	0.0343	0.0077	0.76	0.43
	52	-1.6013	0.0121	0.93	0.31					66	0.0552	0.0108	0.76	0.46
	7	-1.5551	0.0119	0.93	0.42					56	0.0592	0.0108	0.75	0.46
	40	-1.1545	0.0104	0.90	0.41					64	0.0690	0.0108	0.75	0.33
	19	-1.1471	0.0104	0.90	0.34					2	0.0742	0.0076	0.75	0.45
	24	-0.8167	0.0094	0.87	0.42					34	0.1270	0.0075	0.74	0.43
	16	-0.7733	0.0093	0.86	0.36					48	0.1477	0.0075	0.74	0.50
	26	-0.7704	0.0092	0.86	0.45					29	0.1557	0.0075	0.74	0.35
*	3	-0.7694	0.0130	0.86	0.36					36	0.2164	0.0074	0.73	0.29
*	65	-0.6681	0.0127	0.85	0.34					13	0.3362	0.0073	0.71	0.48
*	25	-0.6294	0.0125	0.85	0.42					50	0.3397	0.0073	0.71	0.31
*	18	-0.6121	0.0125	0.84	0.35					37	0.5318	0.0071	0.68	0.30
	63	-0.6074	0.0125	0.84	0.41					17	0.5685	0.0071	0.67	0.31
	35	-0.5999	0.0088	0.84	0.48					14	0.6257	0.0070	0.66	0.24
	10	-0.5688	0.0088	0.84	0.42					11	0.6835	0.0070	0.65	0.39
	41	-0.5402	0.0087	0.84	0.50					58	0.7021	0.0099	0.65	0.47
	6	-0.5379	0.0087	0.84	0.39					54	0.7120	0.0070	0.65	0.33
	39	-0.5181	0.0086	0.83	0.45					68	0.7590	0.0098	0.64	0.45
	32	-0.4953	0.0086	0.83	0.41					42	0.7891	0.0069	0.63	0.28
	53	-0.4323	0.0085	0.82	0.43					28	0.8743	0.0097	0.61	0.39
	49	-0.4208	0.0084	0.82	0.38				#	57	0.9168	0.0097	0.61	0.39
	27	-0.4147	0.0084	0.82	0.44					33	0.9462	0.0069	0.60	0.30
	45	-0.3518	0.0083	0.81	0.42					30	0.9559	0.0068	0.60	0.53
	55	-0.3518	0.0083	0.81	0.43					44	1.0080	0.0068	0.59	0.30
	12	-0.3109	0.0116	0.81	0.35					67	1.0601	0.0096	0.58	0.43
	20	-0.3043	0.0082	0.81	0.30					8	1.0943	0.0068	0.57	0.35
	22	-0.2425	0.0081	0.80	0.46					5	1.1697	0.0068	0.56	0.27
	43	-0.2216	0.0081	0.80	0.36				#	59	1.1776	0.0096	0.56	0.32
	47	-0.1346	0.0079	0.78	0.38				#	69	1.2629	0.0096	0.54	0.30
	62	-0.1313	0.0112	0.78	0.25				#	60	1.3958	0.0095	0.51	0.23
	4	-0.1137	0.0079	0.78	0.49					51	1.5116	0.0067	0.49	0.32
	38	-0.1112	0.0079	0.78	0.28					21	1.6783	0.0068	0.46	0.32
	46	-0.0955	0.0079	0.78	0.39				#	70	1.6935	0.0096	0.46	0.27
	31	-0.0368	0.0078	0.77	0.52					15	2.2840	0.0071	0.35	0.16

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

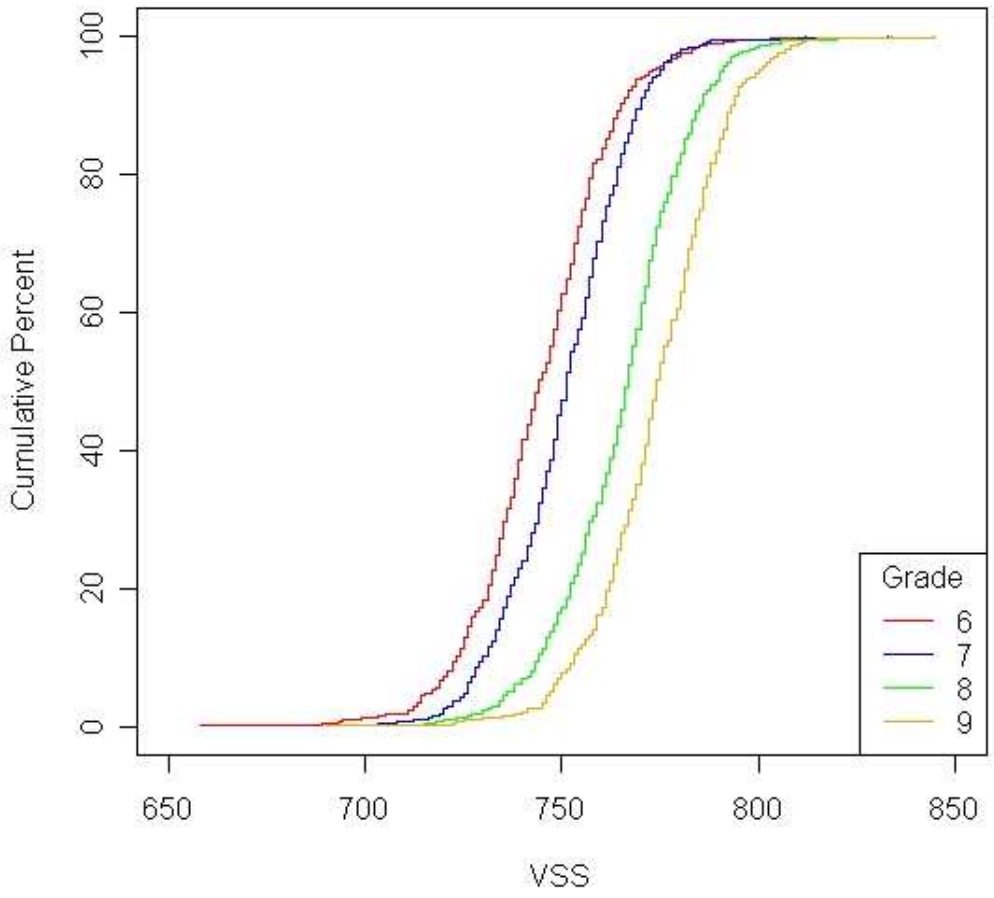
Box plots and relative cumulative frequency (rcf) distributions (illustrated in the following figures) help the researcher visualize the change in year-to-year growth and variability across grades.

Sample Box Plots illustrating
Vertical Scale Distributions Across Adjoining Grades



PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

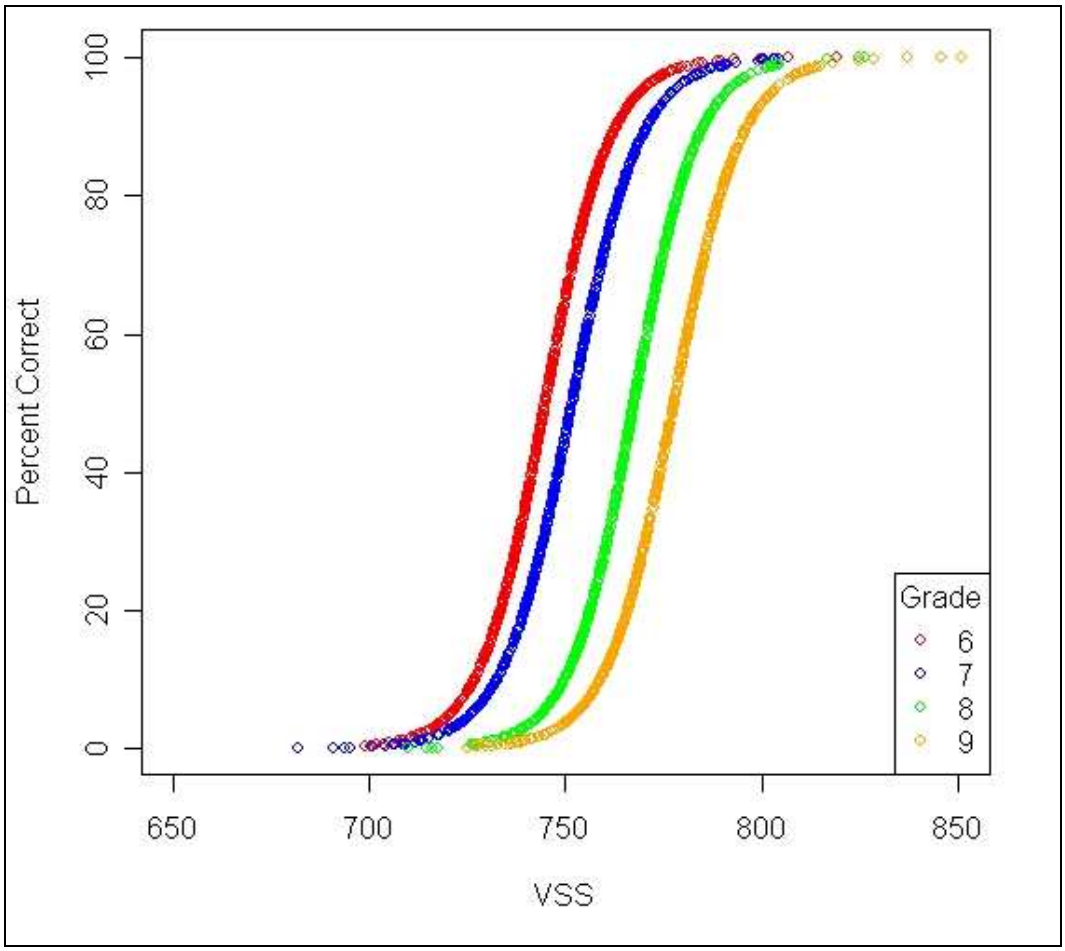
Sample RCFs Illustrating
Vertical Scale Distributions Across Adjoining Grades



PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

In the next figure, Diagnostic Assessment Tool percent correct scores can be plotted on their associated VSS values, which will help illustrate the relative difficulty of each grade-level Diagnostic Assessment Tool test. These are essentially test characteristic curves for each grade-level test represented relative to the vertical scale. Invariably the final results of the scaling study are also tabled, as illustrated in the table below.

Sample Characteristic Curves Across Adjoining Grades



PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Sample Raw-to-VSS Conversion Table

Raw Score	Grade 6	Grade 7	Grade 8	Grade 9
0	573	592	589	604
1	608	626	624	639
2	629	647	645	659
3	642	660	658	672
4	652	669	667	681
5	660	676	675	689
6	666	682	681	695
7	672	688	687	700
8	677	692	692	705
9	681	697	696	709
10	685	701	701	713
11	689	704	705	717
12	693	708	708	721
13	697	711	712	724
14	700	714	715	727
15	703	717	718	731
16	706	720	721	733
17	709	723	724	736
18	712	726	727	739
19	715	729	730	742
20	718	731	733	745
21	720	734	736	747
22	723	736	738	750
23	726	739	741	753
24	729	742	744	755
25	731	744	746	758
26	734	747	749	760
27	736	749	752	763
28	739	752	754	766
29	742	754	757	768
30	745	757	760	771

Correlation and Regression

Correlation analysis is a commonly used measure of the strength and direction of the relationship between two sets of variables and is often used as an indication of predictive validity. Closely related to correlation analysis is regression analysis, which is a collective name for a set of techniques for the modeling and analysis of numerical data consisting of values of a dependent variable (GCA performance) and of one or more independent variables (Diagnostic Assessment Tool performance, additional student-specific variables).

Parsimonious correlation- and regression-based analyses will be used to collect evidence toward the predictive capability of the Diagnostic Assessments. On-grade correlations between Diagnostic Assessment Tool performance and

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

respective GCA subject area performance will be first calculated in Summer 2011 (for Wave 1 courses) and Summer 2012 (for Wave 2 courses). It is only after the first full operational administration of the GCA exams (Spring 2011 for Wave 1 courses and Spring 2012 for Wave 2 courses) that the link between the Diagnostic Assessment Tool and GCA exams can be performed. Classification accuracy analyses will also be conducted to subsequently determine the ability of the Diagnostic Assessment Tool scores to accurately identify individuals who are likely to perform at proficient levels on a given GCA test.

Correlation and regression are the central methodologies behind many predictive validity studies and are significant parts of DRC's research plan. Researchers from a variety of backgrounds are familiar with these procedures, which breeds confidence in their interpretation of regression results. Yet, it is important to emphasize that the same may not be true of many classroom educators, who would prefer to have the study results provided in a more concrete form. Below is a sample of how results might be documented more concretely for the classroom educator. If the link between Diagnostic Assessment Tool and GCA was accomplished with regression, standard errors of estimate could (and should) be provided.

Sample Diagnostic Assessment Tool-to-GCA Conversion Table with Standard Errors

Diagnostic Assessment Tool	GCA		Diagnostic Assessment Tool	GCA		Diagnostic Assessment Tool	GCA	
SS	SS	SE	SS	SS	SE	SS	SS	SE
131	457	29	176	603	29	221	668	29
132	457	30	177	604	30	222	670	30
133	457	31	178	606	31	223	671	31
134	457	32	179	607	32	224	672	32
135	457	29	180	609	29	225	674	29
136	462	30	181	611	30	226	676	30
137	481	31	182	612	31	227	677	31
138	487	32	183	613	32	228	678	32
139	489	29	184	615	29	229	679	29

Having a predicted GCA score and an associated standard error will provide important information to educators. An additional element in the form of providing the *probability of passing* (POP) the GCA will bolster the value of the proposed prediction studies. These probabilities can be derived via the proposed regression methodology and the Rasch measurement model. Specifically, the differences between observed and predicted scores could be divided by the associated standard error of estimate. A sample table that could be provided for any given grade and subject area Diagnostic Assessment Tool is provided below.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Strictly speaking, one could cut off the tabled values at 5 and 95, or 10 and 90, to avoid over interpretations of the precision of results.

Sample Table Showing
Chances in 100 that Students with a Predicted GCA
Score will Equal or Exceed a Given Scaled Score on the GCA

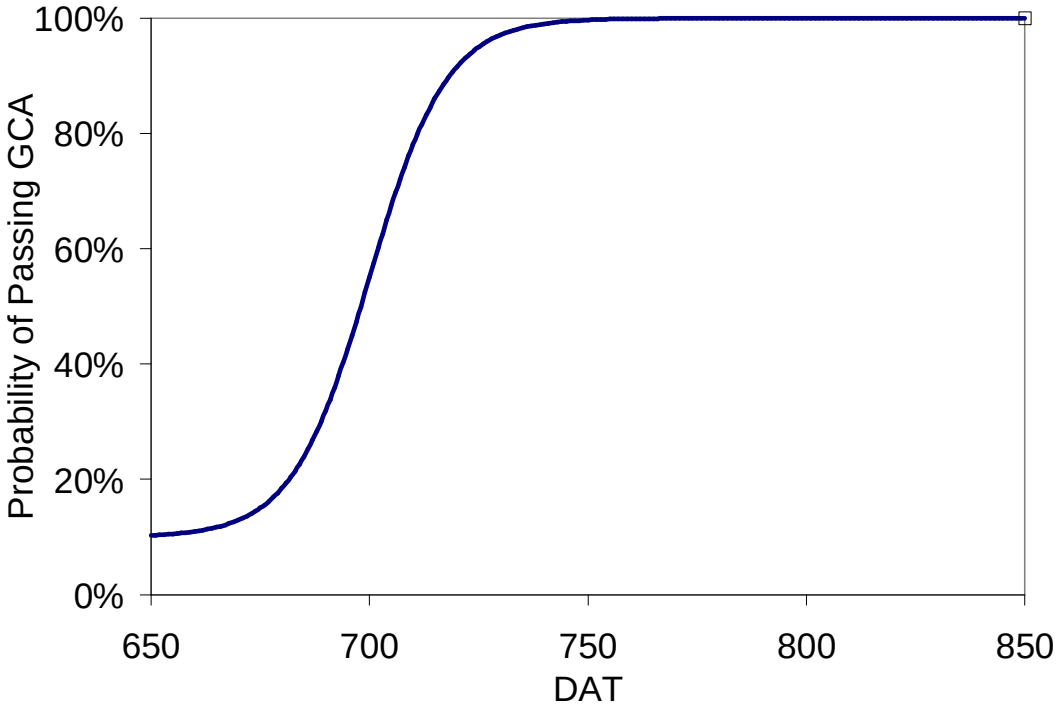
Pred SS*	Actual GCA Scaled Score												
	400	425	450	475	500	525	550	575	600	625	650	675	700
650									99	97	94	87	78
645								99	97	94	87	78	65
640							99	97	94	87	78	65	50
635						99	97	94	87	78	65	50	35
630					99	97	94	87	78	65	50	35	22
620				99	97	94	87	78	65	50	35	22	13
615			99	97	94	87	78	65	50	35	22	13	6
610		99	97	94	87	78	65	50	35	22	13	6	3
605	99	97	94	87	78	65	50	35	22	13	6	3	1
600	97	94	87	78	65	50	35	22	13	6	3	1	
595	94	87	78	65	50	35	22	13	6	3	1		
590	87	78	65	50	35	22	13	6	3	1			

* Predicted GCA SSs derived from regression of actual GCA scores on Diagnostic Assessment Tool performance.

The table above requires certain model assumptions to hold. Actual *probability of passing* (POP) curves can be derived, as illustrated in the next figure. The former tables would be produced during model development and the later developed during cross-validation studies.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

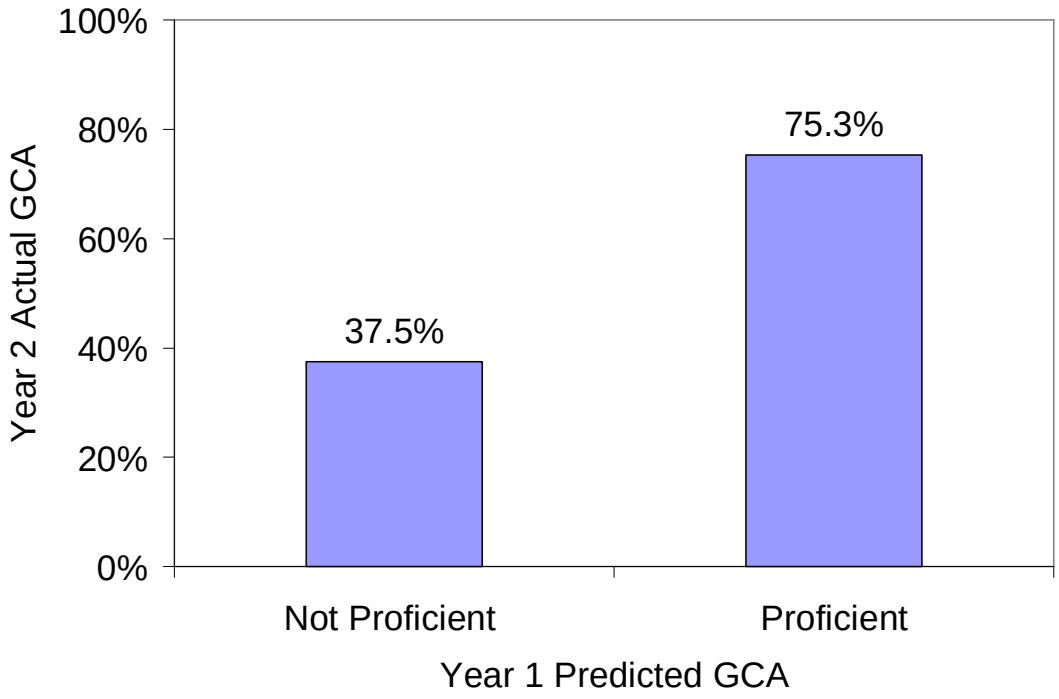
Sample POP Curve



The table and figure above focus on the accuracy of the conditional predictions. The accuracy of the marginal predictions may also be of interest. Here one could plot the percent of those who actually passed the GCAs relative to the dichotomous prediction outcome. For any subject area GCA test, this could be illustrated as shown in the next figure. In the example, 75.3% of students predicted to pass the GCA based on their Diagnostic Assessment Tool score actually passed the GCA (meaning 24.7% "false positive" students did not). On the other hand, there were 37.5% "false negatives" (students predicted to be *Not Proficient* who actually passed the GCA).

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Marginal Prediction Accuracy
Actual GCA Passing Rate Plotted on Predicted GCA Outcome



A comprehensive program for predictive validity requires a long-term investment. This is particularly true for the correlation and regression studies. As an example, regression formulas will require cross validation, and if needed, prediction equations adjusted to improve the model. Improvement could come in several forms: increasing 'hit' rates, minimizing 'false positive' and 'false negatives,' as well as improving 'sensitivity' and 'specificity.'

Linking

In psychometric nomenclature, the term *linking* generally means "putting the scores from two or more tests on the same scale" (National Research Council, 1999, p.15). Projection, a special unidirectional form of linking, can be used to predict or "project" scores on one test from scores on another test without any expectation that exactly the same things are being measured (National Research Council, 1999, p.19). The typical analysis method applied is statistical regression—deriving the predictive distribution for test X performance, given test Y observation. The regression equation can also be conditioned on additional information about the students. Projection is dependent on the way the regression is conducted: specifically projecting test A onto test B gives a different result from projecting test B onto test A (Mislevy, 1992).

Correlations between the two sets of tests—Diagnostic Assessment Tool and GCA— will inform DRC and PDE about the direction and degree of the relationship between the matched tests (e.g., the grade 8 math tests can be linked to the Algebra I GCA). One check of the reasonableness of the relationships

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

would be to study the pattern of correlations between all Diagnostic Assessment Tool and GCA subject area tests (e.g., do the matched tests show higher correlations with each other than they show with other tests?).

Regression procedures will minimize the aggregate vertical distance from the regression line (or curve) to the data points. This distance is the amount of deviation there is when using the predicted score as the estimate of the criterion variable rather than the actual observed score. When using regression procedures, the distributional properties of the predicted scores (e.g., the variance and skewness) will not be the same as the distributional properties of the observed scores. This is acceptable for the Diagnostic Assessment Tool/GCA linking as our interest will always be in the best unidirectional prediction, using Diagnostic Assessment Tool to predict GCA (there is no interest in prediction in the opposite direction). Thus, a symmetric projection relationship is not required.

Standard Errors

A variety of supplemental statistics can be reported to enhance one's understanding of the relationship between linked scores. A strongly recommended approach is to report estimated standard errors for the linking analysis. Standard errors help capture the degree of random fluctuation that one might expect to observe if the linking study were to be replicated. Often the standard errors can be derived theoretically. At other times, a "resampling" procedure such as the bootstrap can be used.

Other Studies

Regression is a simple and direct procedure for linking Diagnostic Assessment Tool performance to the GCA in such a manner to ensure optimum prediction. This research angle also lends itself to addressing the change in the probability of passing the GCAs associated with each point change on the Diagnostic Assessment Tool. More complex analyses could, and perhaps should, be conducted. One methodology in particular seems to hold great promise, hierarchical linear modeling (HLM).

HLM can estimate the effects of students nested within schools. Nesting can create some interpretative challenges if not taken into account. For example, schools can provide more information because they might vary in quality and instructional practices. Ordinary regression will not account for any mediating or moderating effects of nesting variables on individual student variables. In short, the research questions are complex and the structure of the data is complex. HLM is an appropriate technique in this situation.

Additional analyses will include multiple correlations and multiple predictor variables such as longitudinal coefficients (Diagnostic Assessment performance over time), additional academic indicators such as attendance, school-based measures, and special education classification. DRC suggests that demographic variables such as gender or ethnicity would not be included as these variables may be potentially biased and thus artificially inflate prediction coefficients.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

In summary, the broader predictive validity research program for the Diagnostic Assessment Tool may wish to demonstrate that Diagnostic Assessment Tool results are a significant predictor of the GCAs even after controlling for other school and student characteristics that are known correlates with achievement. Ideally, Diagnostic Assessment Tool will be a significant predictor of the GCA and generally indifferent to demographic variables such as ethnicity and family SES (which are typically significant achievement correlates). Only the appropriate analyses will clarify this hypothesis.

VIII.C. ASSESSMENT DELIVERY

VIII.C.1. Online Assessment Delivery System

The CAL online assessment delivery system proposed for Pennsylvania's Diagnostic Assessment Tool is identical to the system proposed for the GCA component, described in detail under *Subheading IV.C.10*. The exact same system features and capabilities will be deployed for both the GCA system and the Diagnostic Assessment Tool, with the exception of the GCA being a linear CBT and the Diagnostic Assessment being a stage-adaptive fixed-length test. Please refer to *Subheading IV.C.10* for a comprehensive description of the CAL online assessment solution for the Pennsylvania GCAs.

VIII.C.2. Supporting Accommodated Assessment

One of the primary goals of the CAL test delivery system is to provide an application that is universally accessible by all students, without the need for adaptation or specialized design for various populations. CAL's testing engine has been designed from the beginning to be accessible and valid with respect to the widest possible range of students, including students with disabilities and students with limited English proficiency. Therefore, the CAL assessment system fully supports the principles of Universal Design, namely equitable use by people with diverse abilities, wide ranging individual preferences and abilities, and tolerance to accidental or unintended actions.

In every state in which CAL has provided online testing solutions we have delivered assessments to special populations and students with diverse abilities, without compromising test security and standardization. Section 508 of the Rehabilitation Act of 1973 details the standards to follow to guarantee a universally accessible application, and the National Center on Educational Outcomes (NCEO) has further articulated the principles of Universal Design to be followed to ensure that assessments are accessible to all students. The CAL online assessment system complies with these standards and has been evaluated in terms of Section 508 compliance.

CAL's universally designed online assessment system significantly reduces the need for accommodations, but does not eliminate the need for all of them. CAL's unique test delivery engine increases the variety of accommodations that can be used without threat to the validity of scores. This will result in inclusive

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

accountability measurement for Pennsylvania, and provides instructionally supportive information across the full range of students.

CAL will work closely with PDE to implement accommodations and modifications to further support students with special needs. Standard CAL tools and features capable of delivering accommodations to support students with various needs include:

- A testing application with strong support for Universal Design principles.
- *Presentation of signed text* in American Sign Language or Signed English.
- *Magnification* of text and images for students with visual impairments.
- *Color overlays* that allow the testing interface to invert the contrast to light text on a black background, with individual choice of text color.
- *Font adjustment* feature that allows students to change text font size.
- Accommodations for students with special learning needs including large-print, audio support, enlarged graphics, zoom features, and a guideline tool to assist with the reading of passages.

Further, standard CAL features such as the Guideline tool, the "Click to Enlarge" tool, various navigation options such as the "BACK" and "NEXT" and the status bar, differing layouts for item display, the "tool help" menu, the summary page, the "Mark for Review" tool, the color-based item status coding, and the warning messages when trying to exit an incomplete test all contribute significantly toward a universally accessible test delivery engine.

VIII.C.3. Costs for Audio Delivery of Assessments

CAL's online assessment system offers state-of-the-art text-to-speech synthesizing capabilities for all students, as required by the specific assessment program. CAL's text-to-speech reader allows students to listen to all text as it is displayed on a computer screen. Students can choose to listen to any part of the text multiple times, and while listening, they may visually follow the text on the computer screen through the visual tracking support offered by a text highlighting feature.

Items with audio files are also supported; however, CAL's text-to-speech synthesizers offer a substantial advantage over audio files since they do not require the downloading of the enormous amount of data required for audio files.

For costing purposes, CAL has assumed that five audio licenses will be required per school (to be installed on five computers). Costs related to additional licenses can be provided upon contract award.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.D. ASSESSMENT MATERIALS

VIII.D.1. Online and Paper Versions of the Assessment Administration Manuals

DRC acknowledges that we will be responsible for providing Assessment Administration Manuals in both paper and electronic versions for the Diagnostic Assessment Tool; the electronic versions will be provided in PDF format for posting online. We propose producing a separate Assessment Administration Manual for each content area that covers all grade levels and courses. The Assessment Administration Manuals will provide complete instructions for using and administering the online Diagnostic Assessment Tool. Paper copies of the manuals will be delivered to all grade 6-12 teachers in participating Pennsylvania schools.

Please see *Subheading VII.E.3., Administration Manuals and Assessment Coordinator's Handbook,* for a detailed discussion of our approach to providing accurate, user-friendly Administration Manuals; our goal is to help streamline the process for busy LEA and school staff. *Subheading VII.D., Assessment Materials,* provides a detailed discussion of our materials production processes.

VIII.D.2. Mathematical Formula/Reference Sheets

DRC's Test Development Team will collaborate with PDE to develop mathematical formula sheets as needed for the Diagnostic Assessment Tool. Formula sheets will be available for student access during a test via the CAL online test delivery system. CAL routinely provides tools such as reference sheets as part of their system.

VIII.D.3. PDE Approval of Final Drafts

DRC understands that final drafts of all materials are subject to PDE approval. We are intimately familiar with PDE's preferred approval process from our work on the PSSA, and we will implement any changes to the process as needed for the GCA Program.

Please see *Subheading IV.A.1, IV.B,* and *IV.C.1* for more information regarding DRC's commitment to meeting project timelines. *Subheadings IV.C.7, IV.C.8,* and *VII.D* provide more information regarding our materials production procedures.

VIII.E. ANCILLARY MATERIALS

VIII.E.1. Diagnostic Assessment Handbook

DRC acknowledges that we will be responsible for providing a Diagnostic Assessment Handbook in both paper and online versions for the Diagnostic Assessment Tool. We propose producing one Handbook to encompass all content areas and grades/courses. The Handbook will provide complete directions for the coordinated use of the online Diagnostic Assessment Tool. Please see

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Subheading VII.E.3., Administration Manuals and Assessment Coordinator's Handbook, for a detailed discussion of our approach to providing accurate, user-friendly Assessment Handbooks; our goal is to help streamline the process for busy LEA and school staff. The Handbooks will be delivered to all participating schools that serve grades 6-12. *Subheading VII.D., Assessment Materials*, provides a detailed discussion of our materials production processes and departments.

VIII.F. Delivery of Online and Print Materials

VIII.F.1. and VIII.F.2. Online Registration and Collection of Student Demographic and Program Information

Online registration of students and collection of individual student demographic information from PIMS for use in the Diagnostic Assessment Tool will be handled in the same manner as proposed for the online GCA test system. Please see *Subheading IV.C.10* for a full description of our approach for handling student registration and PIMS data for online test versions.

VIII.F.3. Administration Manual

Please see *Subheading VIII.D.1* for our proposed development of Administration Manuals.

VIII.F.4. Materials Distribution and Delivery Acknowledgement

Accurate packing, shipping, and delivery of testing related materials are critical for districts and schools to successfully use and administer assessments. DRC is proud of our 16 years of accurate and on-time deliveries for the PSSA. We are committed to bringing that level of service to the GCA Program, including the Diagnostic Assessment Tool component.

For the Diagnostic Assessment Tool, the only materials to be printed and shipped to schools and districts are the manuals and handbooks as described above. They will be distributed at the beginning of each school year for use with the Diagnostic Assessment Tool throughout the year. Please see *Subheading VII.F., Materials Delivery*, for additional information regarding DRC's **ISO 9001:2000-certified** materials packaging, shipping, and delivery processes.

VIII.G. Assessment Administration Supports

VIII.G.1. Customer Service Support

Throughout the entire contract period, LEA, school, and PDE personnel will have the complete support of DRC's and CAL's courteous, professional, "live" Customer Service team to assist with all aspects of the GCA Program, including the Diagnostic Assessment Tool component. Our customer service function is organized such that only staff trained in all components of the GCA Program will respond to calls. The toll-free number and email address will be prominently listed in the Assessment Administration Manuals, the Assessment Coordinator's

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Handbook, and other correspondence with schools and LEAs. Please see *Subheading VII.G.1., Customer Service Support,* for more information on our proposed Customer Service function and process.

VIII.G.2. Procedures for Maintaining Security of Assessment Materials

DRC and CAL are industry leaders in successfully delivering high stakes, large-scale assessments, including the PSSA and other programs in Alaska, Idaho, Louisiana, Ohio, Oklahoma, South Carolina, and Washington. We have proven quality-control and security procedures integrated throughout all of our operational processes. PDE can be assured that all Diagnostic Assessment Tool assessment materials and student responses and data will be handled and stored in a secure manner, as DRC has done over the past 16 years in managing the PSSA.

DRC and CAL understand that ensuring security is critical to maintaining the technical quality, perceived fairness, and integrity of any testing program. We recognize that assessment security is of the utmost importance to PDE.

Please refer to *Subheading IV.C.15.* for detailed information on DRC's security features and procedures, and *Subheading IV.C.10* for a description of CAL's security procedures.

VIII.H. SCORING

VIII.H.1. Scoring Software

CAL's online assessment system captures and archives all online student responses. As described in S*ubheading IV.C.10*, student responses are saved in the event of unexpected interruptions in internet service. If a student's workstation is unable to transmit responses to the testing server during a test due to interruptions in internet service, the responses are saved to an encrypted backup file so the student may either continue testing or exit the system without losing individual response data. Once the network connection is established, the student's saved responses are uploaded to the testing server and the student response file is automatically erased.

The scoring software needed to score, merge, and report student results from the Diagnostic Assessment Tool is inherent to the CAL testing system. Routines are built into the system to score each item as incorrect or correct, to compute a raw score, and to transform that raw score to a scale score. And when the CAT engine is engaged, routines are built into the system to compute ability scores (i.e., theta values) and transform those to scaled scores. See *Subheading VIII.H.3* below for details of CAL's solution for scoring multiple-choice items.

VIII.H.2. Creation of Data Files

CAL will provide all systems and equipment to create, store, and maintain all student-level data and individual student response files in an electronic format in

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

accordance with PDE specifications. Files will contain individual records that are linked to the school district from which they came.

VIII.H.3. Online Real-Time Scoring of Multiple-Choice Items

CAL's solution for the Diagnostic Assessment Tool will provide student scores as soon as a test is finished. Real-time system scoring will be accomplished because all items in the on-line Diagnostic Assessment Tool are multiple-choice items, which can be readily scored by the CAL system. CAL quality assurance specialists will ensure that student responses have passed the quality control steps for key checking and are uploaded accurately to the CAL system along with all other item information. Additional key verification will be conducted after the first 1500 student responses are gathered for each item in the respective grade level and content area item banks. Percentages of students responding to each answer option will be computed in addition to distractor analyses for each item. The proportion of students answering respective items correctly will be verified against available item data (field test results, for example).

DRC Psychometric Services will also conduct key verification analyses. These analyses will be conducted on an early sample of the student tested. Traditional key verification statistics (e.g., p-values, proportion endorsing each response option, point-biserial correlations) will be conducted. See *Subheading VII.J* for further details of DRC's key verification procedure.

VIII.H.4. Storage, Retrieval, and Destruction of Data Files

CAL will store all assessment data throughout the life of the contract in a well-managed, orderly, and quickly retrievable electronic file. Upon request, CAL will provide PDE with access to data files immediately. CAL will destroy all data files contained in the database upon delivery confirmation of written request. If PDE would like a copy of any prior to destruction of files, CAL will provide these in an electronic format.

Rescore requests will be accepted for any Diagnostic Assessment Tool test. In the unlikely event that any assessments or score reports have been inaccurately processed, CAL will reprocess them without additional cost to the requesting party.

VIII.I. Analysis of Assessment Data

Analysis of assessment data will be completed for the Diagnostic Assessment Tool in the same manner as the GCA assessment. Please see *Subheading VII.J* for a detailed look at the analyses.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.J. REPORTS AND DATA FILES

VIII.J.1. Suite of Reports

A key aspect of computerized testing is the immediate return of student scores as students complete their assessments. DRC and CAL understand the activities and coordination required to accurately and comprehensively report results from online test administrations. We believe that the CAL customized reporting system will meet the needs of PDE and Pennsylvania stakeholders as it is not constrained by "off-the-shelf" or pre-packaged products.

CAL has significant experience in reporting activities within the context of large-scale assessment programs. Our extensive experience in diagnostic, formative, and summative assessments, such as those for Alaska, Idaho, Oklahoma, South Carolina, and Kansas, gives CAL the ability to report accurate results immediately upon student completion of an online assessment. Report designs and content are useful and easy to understand. Reports are produced and available to each district/school within critically prescribed time limits. We offer the combination of proven excellence in designing and implementing customized solutions to meet expectations; in-depth understanding of the complexities of assessment reporting; and a cadre of highly qualified professionals who are experienced and have worked collaboratively with our clients to address all reporting requirements, as well as the needs of students, parents, and educators. CAL will also work with DRC psychometric staff and their reporting experts to create a custom reporting solution that is best suited for the Diagnostic Assessment Tool.

CAL and DRC are committed to providing PDE with an innovative reporting package that can be customized to meet the needs of the assessment program. Highlights of CAL's reporting package include the following:

- A comprehensive web-based report delivery system— ResultReporter— that provides for the immediate release of results.
- Reports that are user-friendly, delivered on time, and customizable to meet the evolving needs of Pennsylvania teachers, schools, and districts. They are fully aligned with PDE's desire for various reporting categories and include easily understood data presentations.
- A report process that is established, efficient, and provides high-quality reports.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

ResultReporter, CAL's web-based reporting interface, allows teachers to create customized reports by selecting from a menu of parameters including report types, student lists, tests, time periods, and demographic classifications. The report interface is intuitive, interactive, and diagnostic, and when combined with the TestBuilder interface, will provide the teacher with a flexible classroom assessment system and diagnostic tool.

CAL's ResultReporter provides immediate, real-time results for the following:

- Individual and Aggregate Listings for Item, Part, Total, & Growth Scores, and Classification (proficiency designation)
- Descriptive data summaries by student or user-defined groups such as respective Pennsylvania districts, schools, grade level, course, classroom, Individual Education Plan (IEP), English as a Second Language/Limited English Proficient (ESL/LEP), Section 504, gender, race, and ethnicity
- Graphics to support interpretation and use
- Student-Problem (S-P) Charting for student learning diagnostics
- Data Analysis Rostering Tool which allows users in real time to create and generate reports of their interests (disaggregation of data)
- Contingency Tables (both user-defined and static reports)
- Archives (i.e., maintain and offer past years' data for trend analysis)

Upon test completion, students will receive immediate performance results, along with feedback and recommendations that highlight students' strengths and weaknesses. Student results and feedback will be given for each test item attempted in a way that could be used to analyze students' progress. Students are eligible to take tests multiple times. Below is an example of the type of immediate feedback given to students following test completion. Teachers can elect to have items reviewed immediately following a test or at a later time.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Sample Figure: Immediate Student Feedback

Report Design

Upon contract award, DRC and CAL will work with PDE to determine the design for the Diagnostic Assessment Tool reports. We have thoroughly reviewed the report requirements contained in the RFP and Appendices A and B related to the GCAs and the Diagnostic Assessment. It is evident that PDE desires a similar reporting structure for both assessments, with certain differences that would be inherent to the particular program (e.g., information related to strengths, weaknesses, and predicted performance on the diagnostic student profiles).

In order to maintain consistency in reporting between the two components, DRC and CAL are proposing that the Diagnostic Assessment Tool reports follow a similar design as those proposed for the GCAs. Please see sample reports included in the GCA section of this work plan, *Subheading VII.K., Reports and Data Files*. All reports would be newly designed for these programs, and specific elements would be included that are relevant to the given program, as desired by PDE. Additional examples of reports and data displays that CAL has already built in their system can be found later in this section. These are not intended to be proposed report designs, but rather options for PDE to consider that can be discussed upon contract award.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Report Mockups

The proposed reporting plan for Pennsylvania's Diagnostic Assessment Tool will be reviewed in detail with PDE upon contract award. All reporting specifications will be documented by CAL, with consultation from DRC experts, and approved by PDE prior to final deployment of reports.

Report mockups are an essential element in the report development process. CAL will create report mockups representative of the exact production reports that will be generated and delivered as part of the diagnostic test system. The sample of mockups will consist of representative combinations of parameters and will be comprised of simulated, but realistic, data elements. The mockups will be in the required report layout, display the appropriate fonts and font sizes, and demonstrate online delivery as well as printing capability.

CAL will follow a development process that provides PDE the opportunity to review, change, and approve all mockups prior to reporting system deployment. The mockups will be reviewed by CAL staff for accuracy and consistency, and to ensure they meet the initial requirements. During the review process, PDE will be able to evaluate the static content and layout of each report to make certain they reflect the format, verbiage, and design required. CAL and DRC will work with PDE throughout the review process to incorporate any changes or modifications.

Due dates for the report mockups will be discussed at the planning meeting and finalized shortly after the meeting. The report mockups will be completed, reviewed, and agreed upon by PDE to ensure the final reports meet the requirements.

Student Level Reports

Upon completion of a test, students will receive immediate feedback, showing their personal test results. This information will be a number correct score (e.g., "Good work, Michael. You answered 21 out of 30 items correctly."). When appropriate scaling and linking has been done, students will also receive a scale score that will estimate their chances of succeeding on the GCA course exams (e.g., "Good work, Michael. You answered 21 out of 30 items correctly and received a scaled score of 2300. This gives you an 80% chance of scoring in the proficient level on the GCA exam.").

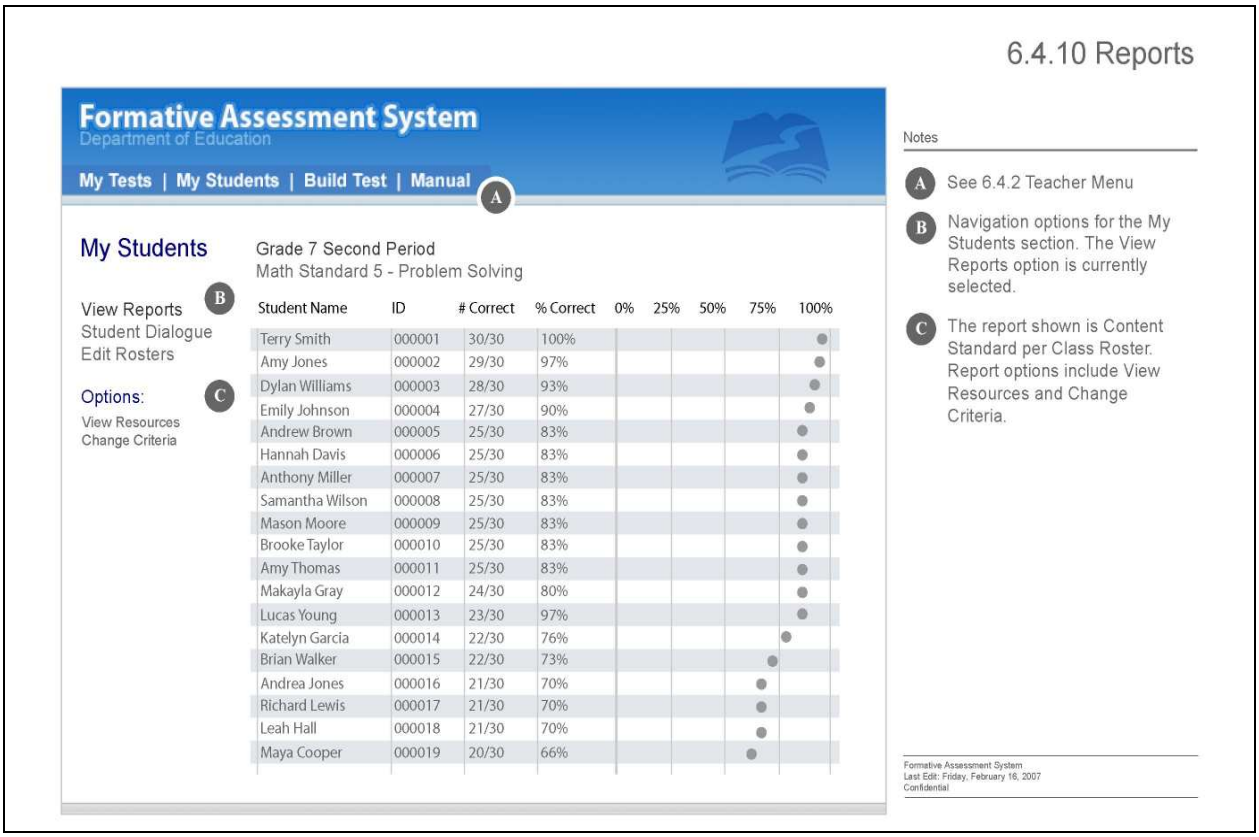
Students will also have the ability to view additional information at a later time. In circumstances where a student has made multiple attempts on the Diagnostic Assessment Tool, the student report will present data relevant to each attempt, such as scale scores. In this way, student improvement from attempt to attempt can be illustrated. This type of emphasis on progress and improvement is indicative of the positive tone that will be typical of student reports. It will be of high priority that feedback given to students is encouraging and constructive in nature.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.J.1.a. Diagnostic Student Profile and VIII.J.1.c. Class Reports

Teachers will have access to online reports that will allow them to review diagnostic student profile and class reports immediately after students complete the assessment. Reports will be viewable online where they may be printed to hard copy or saved electronically if desired. A range of report types will be available to teachers. For example, a Class Report would focus on presenting performance information about how particular groups of students performed on a test. Such groups might include a traditional classroom, a subset of a classroom or classrooms, or even students who match particular demographic indicators like a given gender or ethnicity. Whenever possible, reports in the system will allow teachers to examine data at the most practical levels of granularity.

A sample of this type of roster report is illustrated in the figure below. Again, actual Diagnostic Assessment Tool reports will be designed whenever possible in a style consistent with the GCA report samples provided.



Sample Figure: Roster Report

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Teachers will also have access to reports showing how a student's scale score on the Diagnostic Assessment Tool can be connected to the Model Curriculum. With the stage-adaptive approach described in *Subheading VIII.B., Assessment Construction*, up to five ability groups will be created after Stage 3 testing, with each ability group defined by a range of scale scores. Using an item-mapping type process, content for the Model Curriculum could be "scaled" to match the vertical scale for the Diagnostic Assessment Tool. Students in each ability group could then be directed to the "scale" location in the Model Curriculum for instructional content that best matches their ability level. The table below illustrates how the Diagnostic Assessment Tool results connect to the Model Curriculum resources and materials to enable teachers to find content and interventions that students may need based upon the results. In addition to links to the Model Curriculum units that are matched to the Diagnostic Assessment Tool scale score, sample items can be linked to each scale score range, to show the type of skills that are measured by items at that scale score range.

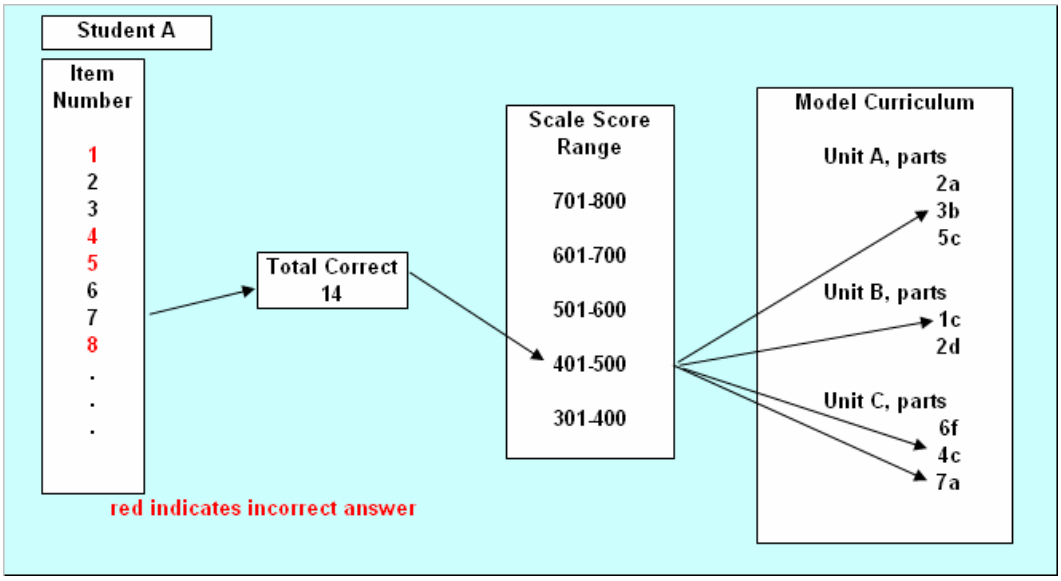


Illustration of How a Student's Scale Score on the Diagnostic Assessment Tool Can Be Connected to the Model Curriculum

In addition, students assigned to each of the five ability groups after Stage 3 will have taken 30 or more items in common—the initial set that all students take in Stage 1, the set that students in each of the three ability groups take in Stage 2, and the set they take in one of the five ability groups in Stage 3. A sample of an item analysis report that could be generated for students in each ability group is provided below.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Click on the question number to view the question.

This indicates the number and percent correct obtained by the student.

STUDENTS ITEMS

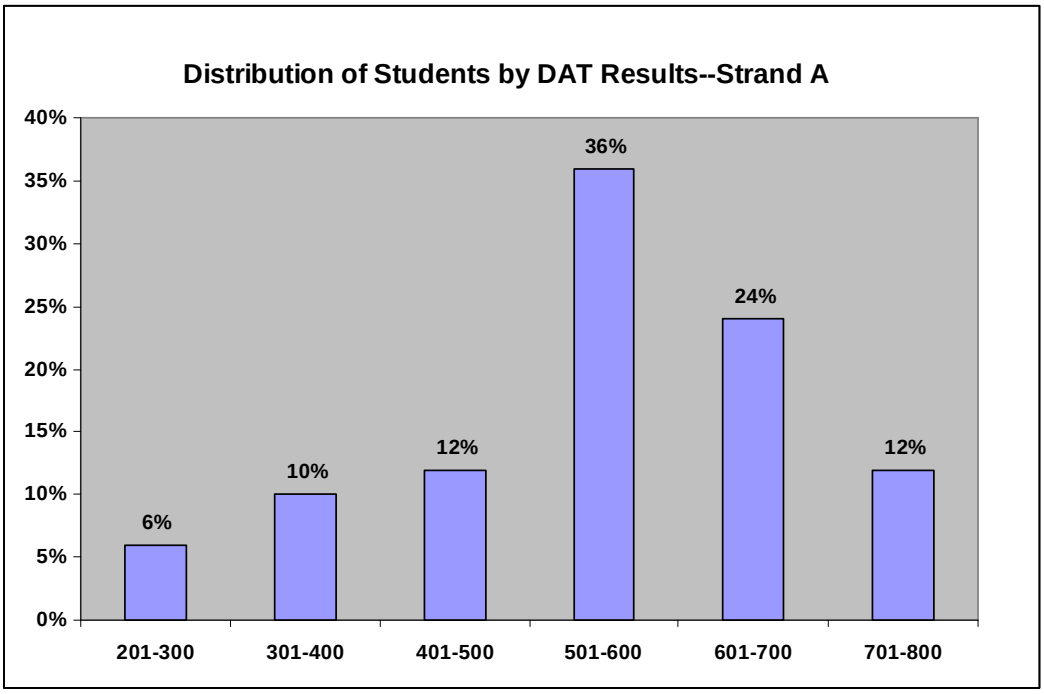
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	score
☐ brad todd	1	B	C	1	C	C	1	1	B	C	1	1	A	1	7/14 **50.0%**
☐ brandy jones	1	1	B	1	1	C	B	1	A	1	1	1	A	C	8/14 **57.14%**
☐ marcia jeans	A	1	1	1	1	1	1	1	1	1	1	1	1	1	13/14 **92.86%**
score	2/3 67%	2/3 67%	1/3 33%	3/3 100%	2/3 67%	1/3 33%	2/3 67%	3/3 100%	1/3 33%	2/3 67%	3/3 100%	3/3 100%	1/3 33%	2/3 67%	

check all :: delete selected :: reactivate selected

A letter indicates an incorrect response. The letter is the answer the student chose.

This indicates the number and percent of students getting this item correct, and indicates how difficult the item is for the group of students.

For the class, teachers could access a report that shows the distribution of their students by scale score range, as illustrated below. Results could be produced for the entire Diagnostic Assessment Tool, or by content "strand."



PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Teachers could use this information to locate resources and materials in the Model Curriculum for students in each scale score range and to facilitate the development of group instructional strategies.

Following are more examples of the types of reports currently provided by the CAL formative assessment system as implemented in Kansas and Alaska.

Formative Test

Report 3: Student-Problem Table with Modified Caution Index explained in a footnote.

Student-Problem (S-P) Table (with Modified

Comprehensive - Includes questions from all assessed indicators

			Easiest.....Hardest													
Student Name	% Correct	MCI	4	8	11	12	1	2	5	7	10	14	3	6	9	13
marcia jeans	92.9%	0	1	1	1	1	A	1	1	1	1	1	1	1	1	1
brandy jones	57.1%	0	1	1	1	1	1	1	1	B	1	C	B	C	A	A
brad todd	50.0%	0	1	1	1	1	1	B	C	1	C	1	C	C	B	A
Percent of students with correct answer:			100	100	100	100	66	66	66	66	66	66	33	33	33	33

See note below for explanation of MCI.

Reviewing the incorrect answers chosen by the students gives information on errors being made or incorrect knowledge

Explanation of Report 3*: Modified Caution Index:* The report above shows how typical a student's correct item responses are in relation to the rest of the students in the same building on the formative test. The modified caution index may range from zero (0) to 99. A student low on the caution index (below 35) answered most of the easier items correctly and most of the difficult items incorrectly. A high MCI (above 35) index may be an indication of guessing, carelessness, or misunderstanding some facet(s) learned by the rest of the students in the building.

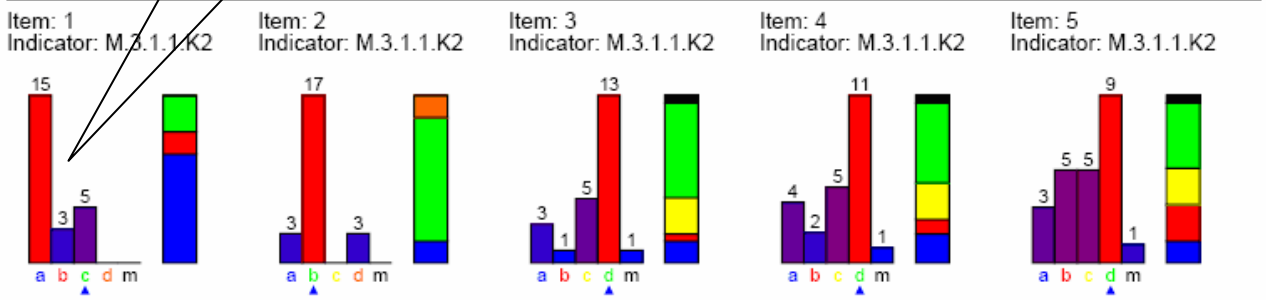
PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

This question posed a problem for the class. The correct answer was "C" but the majority of the students answered "a".

Report 4: Item Performance Summary

Formative Test
Item Performance Summary
3.1.1.k2 - Use numbers to count, order, or denote

District: D0999
Building: CETE Valley
Subject: Mathematics
Grade: 3
Group: All Students (23)



Two bar graphs are shown for each item. The first (clustered) bar graph shows the most common responses for that item in red (tallest bars) and the least common responses (shortest bars) in blue. There is a blue arrowhead that indicates the correct response. The second (stacked) bar graph shows the correct response in green with the A response in blue, the B response in red, the C response in yellow, the D response in orange, and missing responses in black. In both graphs, the height of the bar is directly proportional to the number of responses.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Report 5: Group Summary

Formative Test
Group Summary
3.1.1.k2 - Use numbers to count, order, or denote

District: D0999
Building: CETE Valley
Subject: Mathematics
Grade: 3
Group: All Students (23)

Group Statistics

Number of students 23
Mean 50.7%
Median 55.6%
Mode 55.6%
Standard Deviation 26.3
Range 100.0%
High Score 100.0%
Low Score 0.0%
25th percentile 33.3%
50th percentile 55.6%
75th percentile 66.7%

You can easily tell what indicators were easiest and hardest for your students and adjust instruction accordingly.

Easiest Items

Item #2 (indicator M.3.1.1.K2) had 17 correct responses.
Item #6 (indicator M.3.1.1.K2) had 17 correct responses.
Item #3 (indicator M.3.1.1.K2) had 13 correct responses.

Hardest Items

Item #9 (indicator M.3.1.1.K2) had 10 correct responses.
Item #5 (indicator M.3.1.1.K2) had 9 correct responses.
Item #1 (indicator M.3.1.1.K2) had 5 correct responses.

CAL and DRC look forward to meeting with PDE to design reports for students, parents, and educators.

VIII.J.1.b. Longitudinal Profile

Monitoring students' academic growth over time is an important component of CAL's proposed solution for Pennsylvania's Diagnostic Assessment Tool. The CAL system will provide a variety of longitudinal reporting capabilities to facilitate the monitoring of student progress within grade levels. Reporting of student performance data across years will be provided in eMetric's *Data Interaction*™ dynamic data query and reporting tool.

Longitudinal reporting at grade level can range from a comparison at the content area to the Assessment Anchor level depending on the number and scope of diagnostic tests administered. Four customized longitudinal report formats could be available for teacher and student use.

Single Classroom or Grade-Level Reports

- single parameter
- multiple parameters

Multiple Grade-Level Reports

- single parameter
- multiple parameters

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Data presented in these reports can be filtered by the teacher as described above. CAL will design a reporting system that will minimize the potential that a teacher will use an inappropriate report format for the information assembled for a report. Care is necessary to ensure that "apple and orange" comparisons are avoided using a single parameter report format. All report formats will be submitted to PDE for feedback and approval as part of the report mockup process.

VIII.J.1.d. Building Summary Report
VIII.J.1.e. District Summary Report
VIII.J.1.f. State-Level Report

Building and district reports will be available to all schools and districts on-demand and will represent student results from all students that have completed the respective Diagnostic Assessments. Additionally, state-level reports that include information on usage by school building and aggregate results that provide summary data by district and building will also be available by authorized users at any time.

These reports will be available in eMetric's *Data Interaction*™ dynamic data query and reporting tool.

VIII.J.2. Diagnostic Student Profile Report

Please see *Subheading VIII.J.1.a* for details on the proposed diagnostic student profile report.

VIII.J.3. Accessing Reports Via Secure Website

CAL will engineer a secure website using HTTPS protocols for the authorized access of student reports. Within the CAL system's management application, access to any administrative web interface (such as student enrollment for a test, test-session ticket generation, ticket reactivation, results and reporting, etc.) is controlled by means of administrative accounts.

There can be different levels of administrative accounts (e.g., system, school, teacher), each with different levels of access to functionality and information. Super user accounts will be able to create accounts within/for lower levels of access. Higher-permission managers can restrict (e.g., reports only) or deny access to lower-level account users.

VIII.J.4. Data Query and Reporting Tool

eMetric will provide the same data query and reporting tool for the Diagnostic Assessment as will be used for the GCAs. Data will be securely transferred from the CAL system to eMetric at least three times per year, on a schedule to be determined with PDE. Please see *Subheading VII.K.3* for a full description of eMetric's tools.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.J.5. Sharing Data Files

DRC and CAL will be pleased to work with PDE to supply data to other vendors or educational agencies as needed under this contract. DRC has many years of successful experience working with other vendors/agencies in Pennsylvania under the PSSA contract. This includes receiving data files for the Pennsylvania Alternate System of Assessment (PASA) from the University of Pittsburgh and receipt of data from PIMS for student precode and demographic information, as well as providing assessment results data to organizations such as SAS in Schools and Success for All. We would be happy to work with PDE to share data files as necessary for all components under the GCA Program contract in the same collaborative manner.

VIII.J.6. Providing Data Files to PDE

DRC has successfully defined and produced numerous data files for PDE under the PSSA contract for the past five years. We have worked with PDE to document requirements and define layouts that are intuitive and useful to PDE staff. File layouts are defined while keeping the needs of other vendors and agencies in mind, so that the files can serve multiple purposes.

For the Diagnostic Assessment Tool, DRC will act as an advisor to CAL for file layout definition and data file production. CAL will provide PDE with data files as required by the contract.

VIII.J.7. Transfer of Data Files

CAL understands that PDE will have access to all finalized data files at all times. Upon contract award, we will work with PDE to document the requirements for PDE data file production and dissemination. CAL and DRC staff will ensure that PDE has access to the required data at any point during the life of the contract.

VIII.J.8. Data for Use with Item Samplers

DRC will develop Item and Scoring Samplers each year that include a small number of released items for each content area and grade level/course. We will produce any necessary data to be used as part of the samplers with item level information. DRC has been successfully producing Item and Scoring Samplers for the PSSA for the past four years. Please see *Appendix 3* for sample item and scoring samplers.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

VIII.K. ASSESSMENT SCHEDULE

VIII.K.1. Availability of Diagnostic Tool and Schedule for Development

DRC understands that the Diagnostic Assessment Tool must be operational in Fall 2010 for Mathematics and Fall 2011 for all other content areas. To meet these schedules, we propose the following development and implementation schedule for the Diagnostic Assessment Tool, provided in the table below.

Before becoming operational, the items that are written specifically for the Diagnostic Assessment Tool must be field tested. DRC proposes a Fall 2009 field-test event for the Mathematics component (Grade 6, Grade 7, Pre-Algebra, Algebra I, Algebra II, and Geometry). The field test for Mathematics will allow DRC to develop and try out the items so that they may be operational in Fall (September) 2010.

The field-test event for the Science component (grade 6, grade 7, Biology and Chemistry), the Social Studies component (grade 6, grade 7, U.S. History, World History, and Civics & Government), and for the Language Arts component (grade 6, grade 7grade 8, Literature, and English Composition) will be field tested in Fall 2010. These field-test events will allow DRC to develop and try out the items so that they may be operational in Fall (September) 2011.

Note that there are two operational rollouts of the Diagnostic Assessment Tool (*OP1* and *OP2*). The first rollout coincides with when the Diagnostic Assessment Tool is to be operational per negotiation discussions with PDE. This first rollout of the Diagnostic Assessment Tool will include tools to diagnose the students' strengths and needs of the skills taught in the Model Curriculum. This first rollout does not include a predictive component. To predict the students' chances of success on the GCAs, students must first take the GCAs, which become operational after the *OP1* rollout of the Diagnostic Assessment Tool. After students have taken both the operational Diagnostic Assessment and the operational GCA, the results from those two assessments will be analyzed together to determine the Diagnostic Assessment Tool's predictive capability. The timeframe for when the linking will occur is designated by *Link/OP2* in the table. See *Subheading VIII.B.3.* for details on our linking procedure to ensure the predictive validity of the Diagnostic Assessment Tool.

After the Diagnostic Assessment and GCA have been linked, the Diagnostic Assessment Tool will be released in full operational mode. The Diagnostic Assessment Tool will then have the capability to **diagnose** the students' learning of the Model Curriculum and **predict** their probability of success on the GCA. This full operational mode release is designated by *OP2* in the table.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Development and Implementation Schedule for the Diagnostic Assessment Tool

Diagnostic Assessment Tool		Year 1 (PDE FY 08-09)			Year 2 (PDE FY 09-10)			Year 3 (PDE FY 10-11)			Year 4 (PDE FY 11-12)			Years 5 and 6 (PDE FY 12-13 and FY 13-14)			Year 7 (PDE FY 14-15)		
Wave	Subject	School Year 2008-2009			School Year 2009-2010			School Year 2010-2011			School Year 2011-2012			School Years 2012-2013 and 2013-2014			School Year 2014-2015		
		Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer
1	Mathematics: G6-Geometry		System and Item Development		FT	FT Analysis and Development		OP1	OP1	Link/OP2	OP2	OP2	OP2	OP2	OP2	OP2	OP2	OP2	
2	Science: G6-Chemistry Language Arts: G6-Literature Social Studies: G6-Civics & Government		System and Item Development					FT	FT Analysis and Development		OP1	OP1	Link/OP2	OP2	OP2	OP2	OP2	OP2	

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

DRC proposes operational pool size of approximately 1000 items in each content area (or roughly 160 items at each grade/course). There will be item attrition due to poor items statistics and content, so a larger number of items will be field tested than are needed for the operational forms. This is particularly true for items that are passage-based, as they are for the Language Arts content area. The item count for each of the field-test event are provided in the table below by content area.

Content Area	Targeted Operational CAT Pool Size	Fall 2009 Items to FT	Fall 2010 Items to FT
Mathematics	1,000	1560	---
Language Arts	1,000	---	2,150
Science	1,000	---	1,560
Social Studies	1,000	---	1,560
TOTAL	**4,000**	**1,560**	**5,270**

Field Test Plan

The Diagnostic Assessment Tool items will be field tested as exams, similar in length to the proposed operational Diagnostic Assessment Tool exam. For example, each student will receive 45 total items (each multiple-choice item is worth one point), 30on-grade items and 15 off-grade items (7-8 in each adjacent grade). For example, when a seventh grader logs into the Diagnostic Assessment Tool system to take the field test, the student will be randomly given 30 items from grade 7, seven items from grade 6, and eight items from grade 8. When there is only one adjacent grade (e.g., the highest grade tested), students will take 15 items from that grade or course and ten items from the lower adjacent grade or course. For example, a Chemistry student would take 30 items from the Chemistry course and 15 items from the lower courses or grades (likely Biology and Introductory Physical Science).

The linking of the items across the grades will allow creation of a single (vertical or developmental) scale for the items in each item bank. Thus, scales for Language Arts, Mathematics, Science, and Social Studies will be created. Placing all items across the grade/course span on this one vertical scale will allow the students' scores from the lower grades to be used to diagnose whether the student has acquired the prerequisite skills necessary to perform well at the higher grade or courses. See *Subheading VIII.B.3* for specifics of DRC's linking and equating plan for the Diagnostic Assessment Tool.

DRC's field testing plan calls for participation from selected schools and districts. The representative sample necessary for the field test will be selected from across the Commonwealth. The goal is to have approximately 1000 students take each field test item. Given the number the items to field test (*i*) and the desired number of students per item (*n*), and the number of items (*m*) seen by each student per

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

exam, the total number of students (*N*) needed to field test each content area can be found by using the following equation: $N = (i * n) / m$. The number of items and the number of students needed for each field test event are shown in the table below by content area. Note that the student counts below are based on getting approximately 1000 students per item. The number of students is aggregated across the grade and courses for each content area. DRC looks forward to working with PDE to get the necessary participation from the schools and districts.

Content Area	Fall 2009		Fall 2010	
	Items	Students	Items	Students
Mathematics	1,560	35,000	---	---
Language Arts	---	---	2,150	48,000
Science	---	---	1,560	35,000
Social Studies	---	---	1,560	35,000
TOTAL	**1,560**	**35,000**	**5,270**	**118,000**

Training

In cooperation with PDE, DRC and CAL will develop, plan, and conduct training workshops on the proper use of the Diagnostic Assessment Tool throughout Pennsylvania in accordance with PDE's requirements and approved schedule. We will provide these workshops for LEA and school/testing site test coordinators on an annual basis in preparation for the Diagnostic Assessment administrations. Workshops will be conducted in person at several regional locations around the Commonwealth.

Additionally, DRC will take direction from PDE regarding critical issues and effective methods for the training sessions. DRC has extensive experience providing training for the PSSA to the LEAs and site test coordinators in Pennsylvania for the past five years and provides similar training for many other large-scale assessment clients. DRC and CAL look forward to working with PDE to develop the training and preparation processes for the Diagnostic Assessment Tool. Please see *Subheading VII.G.2* for a full description of DRC's approach to training.

VIII.L. Professional Development

VIII.L.1. Professional Development Plan

CAL's online assessment system for the Diagnostic Assessment Tool is designed to be user friendly for easy access and navigation. To ensure a positive user experience, CAL will also offer video and text formats that teachers can access to obtain professional development resources online in support of the program. These resources will be an integral part of the Pennsylvania's Diagnostic

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Assessment Tool and can be used to learn how to navigate, access, assign, view, and use the system, as well as how to use the Diagnostic Assessment Tool reports to improve instruction and student achievement. These online resources allow educators to access training on the system, reports, and instructional strategies as needed, modeling and describing how each feature of the CAL system works. As an added benefit to increase awareness of the Diagnostic Assessment Tool, DRC and CAL will hold several in person regional workshops each year to strengthen and underscore central understandings of instructional supportive assessment.

There has been extensive research in the area of professional development, which has led to the conclusion that for professional development to be effective it should be ongoing and available to teachers, as it is needed. However, anyone who has provided professional development support or guidance understands the difficulty of on demand training. As such, the CAL system has been designed to provide ongoing professional development through its format designed to improve instruction and student achievement. CAL's diagnostic reports are comprehensive, yet intuitive and user-friendly. This CAL best-in-business feature is advantageous to all educators, including teachers who are not experienced in using student assessment results to guide instruction.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

Works Cited

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

Mislevy, R.J. (1992*). Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.

National Research Council. (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, DC: National Academy Press.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com