VII.K. Reports and Data Files

DRC state assessment clients appreciate our ability to tailor reporting solutions to meet their needs, while still maintaining superior quality and timely delivery.

DRC has nearly 30 years of experience in reporting large-scale assessment results. For every project, DRC works with our clients to **customize our reporting process to the unique needs of their assessment**. 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 will work collaboratively to address all reporting requirements, as well as the needs of students, parents, and educators.

DRC History with Pennsylvania Data and Reports

No other testing vendor has as much experience generating data files and reporting results for Pennsylvania assessment programs as DRC. We understand that the GCA reports must be designed to be easily understood by students, parents, teachers, and administrators to sustain and contribute to the integrity of the testing program. Results of statewide tests are extremely sensitive areas of public disclosure, and a tangible and public representation of the character of the program and all staff associated with that program. To this end, reporting results accurately, effectively, and in a timely manner, with close attention paid to the numerous target audiences, is paramount to the longstanding efficacy of a testing program. DRC staff appreciates this critical element of customer service and client commitment; we are proud of our record regarding the accurate and meaningful reporting of PSSA results for the last 16 years.

Our past performance providing results to Pennsylvania is evidence of DRC's ability to design psychometrically sound and instructionally sensitive reports for the GCAs, and to provide data and reports under tight timelines similar to those required by this RFP. In addition, our reporting experience on other assessments, such as those for Alaska, Louisiana, Oklahoma, South Carolina, and Washington can further assure PDE that DRC has the ability to report accurate results in prescribed time limits.

Our Report Offering for the GCAs

At the core of DRC's proposal is our commitment to continue to provide PDE with accurate and on-time delivery of data and reports. **We are pleased to be able to include eMetric as our exclusive reporting partner for the GCA Program.** eMetric has a history of providing innovative technology-based solutions for displaying and managing assessment data, including several years of direct PSSA experience. Together we have developed a superior reporting offering for the GCAs to include the following:

- Hard-copy, full color student reports designed to present assessment results in an easily understood and psychometrically sound manner.
- Online school and district summary reports focused on the importance of both total test and module results.

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

- eMetric's highly functional *Data Interaction™* dynamic data query and reporting tool.
- Processes and systems that have been successfully used by DRC and eMetric for reporting Pennsylvania results for years, and are familiar to PDE and schools/districts throughout the Commonwealth.

Proposed Timelines for Spring Test Results

DRC fully appreciates PDE's desire to administer the GCAs as close to the end of the school year as possible and quickly get student results in the hands of key district and school personnel following test administration. Given the number of items (multiple-choice and open-ended) to be administered operationally each spring, DRC proposes the following plan for delivering results for each GCA test once standards have been set:

- Final results (scores for both multiple-choice and open-ended items) for graduating seniors will be provided as district student data files posted to DRC's secure Report Delivery System website no less than 10 calendar days prior to graduation. DRC understands that graduation dates vary by district and will adhere to a particular district's graduation schedule when providing results.
- Final results (scores for both multiple-choice and open-ended items) for all non-graduating students will be provided to districts before the end of the school year. These results will also be provided as district student data files posted to DRC's secure Report Delivery website.
- Hard copy student reports and electronic summary reports will be provided in a timely manner as outlined under *Subheading VII.M., Assessment Schedule*.

Providing Preliminary Results

Per discussions with PDE, DRC no longer proposes reporting students' multiple-choice-only scores within the 7- or 14-day timeline. We understand that the results for non-graduating students must be in the districts before the end of the school year. DRC will commit to providing these results within this timeframe.

Per discussion with PDE, DRC will provide raw-score-only roster reports for students for those administrations that occur before standard-setting activities. For the Fall (December) 2010 and Spring 2011 administrations of the Wave 1 courses, raw-score-only roster reports will be provided for these students. Final results (including scale scores and performance levels) will be provided after standard-setting activities. Similarly, for the Wave 2 courses, students taking the exams during the Fall (December) 2011 and Spring 2012 administrations will receive raw-score-only roster reports.

Specific details regarding our proposed reporting timelines can be found under *Subheading VII.M., Assessment Schedule*.

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

VII.K.1. Hard Copy and Electronic Reports

VII.K.1.a. Student/Parent Report

DRC is committed to developing reports that can be adjusted to reflect the evolving needs of the GCAs and delivering those reports **on time**. The design of the reports will be **user-friendly** and feature **clear graphics** and **color** to represent various data elements. Additionally, the graphic design elements will be similar to those used on summary reports for a cohesive look and feel.

Our proposed student/parent report will contain the student's name, school, district, test date, and course. It will present summary information in the form of scaled score and performance level achieved for each module tested. Below we have included images of the front and back of our proposed student/parent report.

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

PENNSYLVANIA
Graduation Competency Assessment Student Report

Dear Family,

This report provides you with information on your student's performance on the Graduation Competency Assessments for the course listed.

Use the information in this report to discuss your student's performance with your student's teacher. A strong partnership between families and teachers is critical for your student's success.

Yours truly,

Gerald L. Zahorchak, D.Ed.,
Secretary of Education

Name:	Pennsylvania Student
School:	Pennsylvania School
District:	Pennsylvania District
Test Date:	Spring 2010
Course:	Geometry

Achieving the Goal:
Proficiency for All Students

Module	Performance Level
ONE	Proficient
TWO	Proficient
THREE	Non-Proficient

What are the Graduation Competency Assessments?

To ensure the academic preparedness of Pennsylvania's high school seniors, all Pennsylvania high school students are required to demonstrate a level of proficiency on Pennsylvania's academic standards to graduate. Students can demonstrate proficiency by scoring proficient or advanced on the 11th grade Pennsylvania System of School Assessments or 12th grade retest, or by passing a progressive series of Graduation Competency Assessments in the core academic areas of math, science, English, and social studies.

The Graduation Competency Assessments are comprised of ten content areas: Geometry, Algebra I, Algebra II, Biology, English Composition, US History, World History, Literature, Chemistry, and Civics and Government.

Each content area is divided into three modules that measure a student's knowledge of the content area. A student must pass all three modules to show competency for a content area.

Your student's performance is detailed on the back side of this report.

For more information on the Graduation Competency Assessments, go to www.pde.state.pa.us

The Pennsylvania Graduation Competency Assessment

www.pde.state.pa.us

Front Page of Proposed Student/Parent Report

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

Family Connections

Strengthening ties between the family and school

Student Results for GEOMETRY

Module	Points Possible	Student's Score	Scaled Score	Performance Level
ONE	20	15	850	Proficient
TWO	20	13	800	Proficient
THREE	20	10	745	Non-Proficient

Module One Score: 850 **Performance Level: Proficient**

Non-Proficient | Proficient (750)

Your student's score is indicated by the ↓ above. If your student tested again, his or her score would likely remain in the following range: 825–875.

Module Two Score: 800 **Performance Level: Proficient**

Non-Proficient | Proficient (750)

Your student's score is indicated by the ↓ above. If your student tested again, his or her score would likely remain in the following range: 780–820.

Module Three Score: 745 **Performance Level: Non-Proficient**

Non-Proficient | Proficient (750)

Your student's score is indicated by the ↓ above. If your student tested again, his or her score would likely remain in the following range: 725–765.

Maximizing Mathematics Achievement

One way to measure the efficiency of a 3-dimensional shape is to determine the ratio of surface area to volume. Imagine a container. Surface area is a measure of the amount of material used to build the container. Volume is a measure of the amount of storage space in the container. The student should pick out several storage containers (jars, soda cans, shoe boxes, etc.). For each one, find the ratio of the surface area to the volume. He or she should record the results in a table. Which shape has the smallest ratio of surface area to volume? Which has the largest?

What do the Performance Levels mean?

Proficient: Satisfactory academic performance indicating a solid understanding and adequate display of the skills included in Pennsylvania's Academic Content Standards.
Non-Proficient: Marginal academic performance, work approaching, but not yet reaching, satisfactory performance. Performance indicates a partial understanding and limited display of the skills included in Pennsylvania's Academic Content Standards, and the student may need additional instructional opportunities and/or increased student academic commitment to achieve the Proficient level.

Where can I find more information to assist my child?

The graduation competency assessments are part of a multifaceted program that includes a model curriculum, a diagnostic analysis tool, as well as the graduation competency assessments. The Pennsylvania Department of Education at www.pde.state.pa.us provides resources to help your child improve. Click on the **Pre-K-12** tab and then click on the **Assessment** link. Select from the menu on the left side of the page. One available tool is the *Assessment Handbook*, which is designed to help students prepare for assessments. Also available are the subject-specific item samplers. The samplers include released student items, including actual student constructed responses that have been scored and annotated.

The Pennsylvania Graduation Competency Assessment

www.pde.state.pa.us

Back Page of Proposed Student/Parent Report

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

As shown above, on the back of the full-color, double-sided report, DRC will present detailed information regarding the student's performance on each of the modules tested for a given course. This information will be presented graphically with accompanying explanatory text. DRC's proposed student/parent report features the following data elements for each module tested within a course:

- Scaled score and performance level.
- Standard error of measurement displayed as a scaled score range.
- Total points possible and student's raw score.

To help students and parents understand the meaning of the performance levels, DRC is proposing a section of the report that includes general performance level descriptors. We will also include information regarding resources for students and parents to find more information about the GCAs.

Full-size samples of DRC's proposed GCA student/parent report can be found under the *Reports* tab at the end of this section. This design is subject to PDE approval; we look forward to working with PDE on the design of the GCA student/parent reports upon contract award.

Student/Parent Report Hard-copy Distribution

Hard-copy student/parent reports will be produced using the same high quality secure materials production procedures that DRC has used for years for the PSSA. DRC will perform a thorough quality assurance review prior to the production of data files and reports. All data and reports are thoroughly tested to guarantee accuracy. One hardcopy per student per GCA course will be provided. Additional copies can be printed and shipped if desired. This can be negotiated upon contract award.

Following the printing of the reports, DRC will assemble them by school and district, and package them into boxes clearly labeled "Test Results Enclosed—OPEN IMMEDIATELY." Packaged reports will be shipped directly to districts for distribution to schools, except for districts with 10 or more schools, in which case the reports will be shipped directly to schools.

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

VII.K.1.b. and VII.K.1.c. GCA Summary Report and Comprehensive Summary Report

DRC is proposing school summary and comprehensive district summary reports using a similar look and feel to the student/parent report for a cohesive approach to GCA results. Our design for summary reports incorporates the following features desired by PDE:

- Percentage of students performing at each performance level for each module by course.
- School performance compared to district and state performance on the school summary report, and district performance compared to state performance on the comprehensive district summary report.
- Raw score averages by course, module, and assessment anchor (for each anchor with at least eight included items).

Examples of these data displays on the summary reports are shown in the graphics below.

Percentage of Students Proficient*

Course	School	District	State
Algebra I	69	72	65
Algebra II	67	70	62
Biology	65	65	71
Chemistry	73	76	70
Civics and Government	80	88	82
English Composition	85	80	78
Geometry	70	78	75
Literature	82	80	81
US History	72	75	70
World History	71	75	69

*These percentages reflect students who are proficient for all three modules in a course.

Example from Front Page of School Summary Report
Comparison of Percentage of Students Proficient by Course

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

Pennsylvania School District
2010 Performance Level Distribution by Course and Module

Total Number and Percentages

	Total Number	Average Raw Score	Average Scaled Score	Percent Proficient	Percent Non-Proficient
Algebra I					
Module one	356	17	812	80	20
Module two	356	16	748	70	30
Module three	356	12	684	60	40
Algebra II					
Module one	345	16	799	78	22
Module two	345	14	735	68	32
Module three	345	12	696	62	38
Biology					
Module one	330	15	786	76	24
Module two	330	12	722	66	34
Module three	330	10	684	60	40
Chemistry					
Module one	290	15	780	75	25
Module two	290	13	716	65	35
Module three	290	13	652	55	45
Civics and Government					
Module one	356	18	876	90	10
Module two	356	18	844	85	15
Module three	356	16	812	80	20
English Composition					
Module one	356	16	786	76	24
Module two	356	15	760	72	28
Module three	356	11	735	68	32
Geometry					
Module one	326	12	799	78	22
Module two	326	10	748	70	30
Module three	326	13	716	65	35
Literature					
Module one	356	15	844	85	15
Module two	356	16	824	82	18
Module three	356	12	805	79	21
US History					
Module one	275	13	799	78	22
Module two	275	15	760	72	28
Module three	275	12	748	70	30
World History					
Module one	356	17	812	80	20
Module two	356	18	812	80	20
Module three	356	17	812	80	20

NOTE: Percentages reported are for all grades combined.

District Comprehensive Summary Report
Average Scores by Course and Module

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

Pennsylvania High School
2010 Performance by Course and Reporting Category

Algebra I Reporting Categories and Assessment Anchors

	Total Points Possible	State Average	District Average	School Average
Module One	**20**	**13.2**	**14.8**	**15.3**
M.1.1	8	6.5	6.1	7.1
M.1.2	8	5.8	6.2	5.5
Module Two	**20**	**13.2**	**14.8**	**15.3**
M.2.1	8	6.5	6.1	7.1
M.2.2	8	5.8	6.2	5.5
Module Three	**20**	**13.2**	**14.8**	**15.3**
M.3.1	8	6.5	6.1	7.1
M.3.2	8	5.8	6.2	5.5

Algebra II Reporting Categories and Assessment Anchors

	Total Points Possible	State Average	District Average	School Average
Module One	**20**	**13.2**	**14.8**	**15.3**
M.1.1	8	6.5	6.1	7.1
M.1.2	8	5.8	6.2	5.5
Module Two	**20**	**13.2**	**14.8**	**15.3**
M.2.1	8	6.5	6.1	7.1
M.2.2	8	5.8	6.2	5.5
Module Three	**20**	**13.2**	**14.8**	**15.3**
M.3.1	8	6.5	6.1	7.1
M.3.2	8	5.8	6.2	5.5

Biology Reporting Categories and Assessment Anchors

	Total Points Possible	State Average	District Average	School Average
Module One	**20**	**13.2**	**14.8**	**15.3**
M.1.1	8	6.5	6.1	7.1
M.1.2	8	5.8	6.2	5.5
Module Two	**20**	**13.2**	**14.8**	**15.3**
M.2.1	8	6.5	6.1	7.1
M.2.2	8	5.8	6.2	5.5
Module Three	**20**	**13.2**	**14.8**	**15.3**
M.3.1	8	6.5	6.1	7.1
M.3.2	8	5.8	6.2	5.5

Chemistry Reporting Categories and Assessment Anchors

	Total Points Possible	State Average	District Average	School Average
Module One	**20**	**13.2**	**14.8**	**15.3**
M.1.1	8	6.5	6.1	7.1
M.1.2	8	5.8	6.2	5.5
Module Two	**20**	**13.2**	**14.8**	**15.3**
M.2.1	8	6.5	6.1	7.1
M.2.2	8	5.8	6.2	5.5
Module Three	**20**	**13.2**	**14.8**	**15.3**
M.3.1	8	6.5	6.1	7.1
M.3.2	8	5.8	6.2	5.5

Example from School Summary Report
Averages by Course, Module, and Assessment Anchor

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

Full-size samples of DRC's proposed school and district summary reports can be found under the *Reports* tab located at the end of this section. Upon award, DRC would be pleased to discuss additional data elements to potentially include on the summary reports. For example, on the School Summary Reports, PDE may wish to consider including two-year comparative data for the district and Commonwealth, as well as the school. This could enable the analysis of school-level trends in the context of trends happening within the district or the Commonwealth.

Option for Report Interpretation Guides

As an additional GCA report offering, DRC would like to propose the development and production of Report Interpretation Guides as an option. These guides can be useful tools for schools and districts as they seek to better understand all aspects of their reports and how to use their results effectively. If PDE is interested in this option, costs can be provided upon contract award.

VII.K.2. Accessing Reports via Secure Website

The school and district summary reports will be available from within eMetric's *Data Interaction*™ data query tool and will be provided as pre-defined reports or macros that are printable in full color. As they are provided on the same platform as the dynamic components of the *Data Interaction*™, users can utilize the dynamic features of the data query tool and drill down from the data on these summary reports to individual student results or subgroup results.

The district and school summary reports will be provided online as described below:

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

- **Predefined Reports:** School and district summary reports will be available as options that can be selected from the "Report Selection" menu in the data query tool. These options will be enabled based on the user type. District level users will be able to access the district wide summary report and also access the school summary report for any school within the district. School level users will only be able to access the school summary report for their respective schools. Users can select this option and obtain the comprehensive summary report compiled as a PDF document. This PDF document will be formatted for printing and can be distributed within the school or district.
- **Macros:** Sections of the school and district summary reports will also be made available as macros currently available in the data query tool. Users will be able to select a specific macro from a list, and obtain an online presentation of the school/district summary report. Reports obtained through macros will be interactive and include options to drill down, where appropriate, from the summary report to the individual performance reports. Export options to Excel (CSV) and PDF will also be provided on the reports generated using the macro option.

More information regarding eMetric's *Data Interaction*™ system can be found later in this section under *Subheading VII.K.3*.

VII.K.3. Data Query and Reporting Tool

DRC and eMetric are proud to propose eMetric's *Data Interaction*™ system as our solution for the required data query and reporting tool. PDE has experienced the superior functionality of this tool for the past four years because it has been used for PSSA, PASA, and ACCESS for ELLs assessment data. This is clearly the best solution for the continued, uninterrupted management of Pennsylvania student assessment data, with no transition required. PDE can be assured that users will receive the same high quality product and service from eMetric as they have since 2004.

Data Interaction™ for the GCA Program: Overview

Data Interaction™ for Pennsylvania Student Assessments is a dynamic data query and reporting tool for the dissemination and analysis of assessment results from Pennsylvania statewide assessment programs. It is a secure-access application provided for educators and policy makers at the state, intermediate unit, district, and school levels. It has been in use throughout the Commonwealth since the winter of 2004. eMetric's services include customization, development, load-balanced hosting, maintenance, and technical support for the entire web application.

To meet the requirements specified in the RFP for a dynamic data query and reporting tool for both the GCAs and the Diagnostic Assessment Tool, eMetric proposes to create an enhanced and expanded version of *Data Interaction*™ (please see *Subheading VII.K.3* and *Subheading VIII.J.4)*. The new version will:

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

- Provide an integrative environment where users can access the assessment results from both the GCAs and the Diagnostic Assessments.
- Provide a longitudinal component to track student performance across administrations for the diagnostic assessment and GCAs, including retest administrations.
- Provide a feeder component for users to keep track of performance across administrations based on the current roster.
- Provide a data integration component for users to view GCA assessment results along with PSSA assessment results, where appropriate.

Content Areas and Administrations

Data will be available in the *Data Interaction*™ system after each operational administration of the GCA. GCA Wave 1 data (Algebra I, Algebra II, and Geometry) will be available after the first operational administration in Spring 2011 and after each subsequent administration. GCA Wave 2 data (Biology, Chemistry, Literature, English Composition, US History, World History, and Civics & Government) will be available after the first operational administration in Spring 2012 and after each subsequent administration.

Reporting Components

The *Data Interaction*™ system will provide the following reporting components:

- Group Summary Report
- Graphical Summary Report
- Individual Performance Report
- Longitudinal Roster Report
- Longitudinal Feeder Report

Screenshots illustrating many of these components can be found at the end of this section.

Group Summary Report provides summary statistics including, but not limited to, number tested, mean scale score, number and percent of students that passed or did not pass, average raw scores by anchor, maximum score possible, and percent of score points earned for each anchor. Users can select content areas, statistics, administrations, demographic variables for summary, test format (e.g. paper or online), and different report views. Users can also drill down to subgroup or individual student results (see Figure 1). eMetric would work with DRC and PDE to define the specific scores that would be available on this report.

Graphical Summary Report provides summary graphs including bar charts, pie charts, histograms, and line graphs. The statistics to be graphed include, but are not limited to, percent of students that passed or did not pass by module or by total test, or percent of score points earned for each anchor. Users can select

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

content, statistics, demographic variables, multiple administrations, and test format (e.g. paper or online). Users can also drill down to individual student results (see Figure 2).

Individual Performance Report provides individual student scores including scale scores, pass/did not pass, and total raw scores by anchor. Users would be able to view reports by selecting a GCA or a diagnostic assessment administration. It also provides student level data including local and state ID, demographic information, and score attribution information. Users can select subgroups of students by filtering options, sort data in ascending or descending order, and perform ad hoc queries on any score variable to obtain summary reports or graphs (see Figure 3). eMetric would work with DRC and PDE to determine the scores that would be available on this report. An additional level of drill-down to the online student report can be incorporated upon PDE request. DRC and eMetric will be pleased to provide costs for this option upon contract award.

Longitudinal Roster Report allows users to track student performance from administration to administration at the individual student level. Users may choose to view longitudinal results for GCAs only for retest students, or view longitudinal results on different diagnostic assessments, or view both the diagnostic assessment and GCA results in a single report (see Figure 6). eMetric would work with DRC and PDE to determine the data elements to be made available for this report. It should be noted that this report is limited to student results collected in the same school for school level users or the same district for district level users and is based on the roster of students at the time of test administration.

Longitudinal Feeder Report allows users to track student performance from administration to administration based on the roster of the current academic year or semester and is not confined to the assessment results collected in the same school or district. This report requires a data import from PIMS into the Data Interaction System, containing a statewide roster for the new school period (year, semester, or grading period) that should include the PA secure ID, current district and school numbers and GCA enrollment (see Figure 6). This roster can be imported as often as necessary to keep the feeder report up-to-date.

System Features

System features of the *Data Interaction*™ system include:

- Secure access
- Ad-hoc query interface
- Save and organize queries
- Data analysis tools
- Display features
- Graphics and data visualization

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

- Export options
- Administrative configuration

Secure Access

eMetric's *Data Interaction*™ provides secure access to sensitive student demographic information and achievement data. The system provides role-based security with privileges for each role as defined by the administrator. Data transmissions between users and the server are encrypted. Sensitive data is stored in an encrypted format on the server with the server infrastructure and perimeter secured using industry standard best practices. *Data Interaction*™ is compliant with the security and operational requirements specified in both the FERPA and COPPA policies. The system also provides the necessary infrastructure to handle the password reset requests from schools and districts with a workflow that does not disclose any log-in information to PDE or the Intermediate Unit.

eMetric will provide "level 1" single sign-on capabilities with Ed-HUB. eMetric will create an application programming interface (API), which interfaces with the Ed-HUB portal component to retrieve credentials and the web service to access/retrieve user information from the Education Directory for use in the *Data Interaction*™ system. eMetric will work collaboratively with DRC and designated PDE staff to accomplish the required "level 1" integration.

The existing *Data Interaction*™ user accounts will continue to work until the completion of integration with Ed-HUB. eMetric proposes a system where the existing legacy user accounts will co-exist with the integrated Ed-HUB directory to ensure seamless integration and minimize any user frustration or confusion. To remediate the redundancy of user accounts, the "Control Panel" feature, which is currently available, will be made accessible only for users logging on using legacy user accounts.

Ad-Hoc Query Interface

Data Interaction™ provides the ability to run ad hoc queries on the data in a user-friendly manner. The query interface provides an easy way to select the type of report, data elements to be displayed, and the type of operation (e.g., grouping, search, exclusion, calculation) to be performed. Customization of queries can be accomplished by using either the selection menu or the selection tabs provided on the report window.

Save and Organize Queries

Each query generated by *Data Interaction*™ can be saved for future reference without having to reset parameter selections. Users can create and save unlimited queries and organize them into a folder structure for easy reference and retrieval. Additionally, each saved query can be directly exported to a CSV file without having to load the actual report. Furthermore, users can add annotations or comments to aid in the interpretation of specific queries. These annotations are attached to queries in a non-intrusive manner and provide a means for collaborative interpretation of student performance results (see Figures 1 and 3).

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

Data Analysis Tools

Data Interaction™ offers a wide variety of data analysis features that range from drill-down disaggregations and calculations to advanced functions for uni-variate and bi-variate analyses.

- *Drill-down*: *Data Interaction*™ provides extensive drill-down options for disaggregating student populations by multiple demographic variables, navigating from summary level data to a roster report for a selected subgroup, or drilling down from the roster report to obtain detailed individual student results with a narrative description and diagnosis of the student's performance (see Figures 1, 3, and 4).
- *Calculate Percentages/N*: This feature allows users to convert values such as N counts, raw scores, etc., shown on the reports into percentages by simply clicking on the column header and choosing the "Calculate Percentages" option, or to obtain n-count values from percentages that are reported for various data elements (see Figure 1).
- *Summarize/Summarize by*: This feature allows users to quickly obtain a summary of the student population by clicking on the column header and choosing the "summarize" option. Statistics shown on the summary include Valid N, Total N, Mean, Standard Deviation, Minimum and Maximum. The *Summarize by* option allows users to obtain the same statistics by subgroups (see Figure 3).
- *Correlate*: This feature allows users to compare student performance on different score variables by clicking on the column header of a score variable and choosing the "plot against" option. The data are shown as a scatter plot where score variables are represented as the axis for the graph with a regression line displayed along with prediction bands. Users can then identify outliers and drill down to the roster of individual students. The scatter plot also allows users to zoom in or out from selected regions of the graph and to obtain statistics including n-count and correlation coefficient (see Figures 3 and 4).
- *Distribute/Distribute by*: This feature allows users to obtain the frequency distribution of a specific score variable within the student population by simply clicking on the column header and choosing the *Distribute* or *Distribute by* option. Helpful statistics such as cumulative frequency and percentages are also shown on the distribution report. *Distribute by* allows users to obtain score distributions by subgroups (see Figure 3).
- *Crosstab*: For categorical or ordinal data, *Data Interaction*™ allows users to obtain a two-way tabulation of the data for examining trend and/or association between two variables. Users may view n-count or percentages by cell, row, or column (see Figure 5).

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

Display Features

Data Interaction™ provides a rich set of data display features that include:

- *Hide/Show Columns*: With a single click on the column header and selection of either "show" or "hide," *Data Interaction*™ appropriately includes or deletes the column from view. This flexibility enables users to customize their report display without having to navigate back to the selections screen and select/deselect the data elements for the report.
- *Paging*: The reports generated on the *Data Interaction*™ system usually involve large amounts of data, which would be too cumbersome to display on a single screen. Consequently, eMetric has included a "paging" functionality to display a user-defined number of records on a single "page." Users can navigate between pages by clicking on the appropriate page navigation buttons, jump to a specific page number (see Figures 1 and 3) and select how many records to display per page.
- *Transpose*: With a single click of a button, users may transpose the report table for a different view. The transpose function is available for both the summary report and the individual performance report (see Figure 1).
- *Sorting*: The reports generated on the *Data Interaction*™ system utilize a sort functionality, which offers bi-directional sorting (ascending, descending) by clicking on the appropriate column header and choosing the sort option. Users may sort by a single variable or multiple variables (see Figure 3).

Graphics and Data Visualization

Data Interaction™ provides a variety of data visualization and graphical features, which enable data to be represented in the form of bar graphs, histograms, line graphs, scatter plots, and pie charts. A powerful feature of *Data Interaction*™ is the ability for authorized educators to drill down from the graphs and charts to individual student rosters. Beyond providing users with the group's overall performance, this attribute enables educators to isolate and identify students who are exemplary performers or those who may require special assistance. The graphs and charts can easily be exported from the system by saving them as graphical file formats (JPEG, PNG) or as PDF files to be included in presentations and other reports (see Figure 2).

Export Options

All tabular reports generated using the *Data Interaction*™ system can be exported to a CSV file or PDF reports. Large collections of reports such as Individual Performance Reports for a school can be queued into the system and downloaded after the PDF reports are generated. Graphs and charts generated using the system can be saved as images on a user's local computer or be saved in the PDF format (see Figures 1 and 3).

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

Administrative Configuration

Data Interaction™ provides LEA administrative users the flexibility of managing the inventory of data elements that can be included on the reports using the Administrative Configuration interface to enable/disable specific data elements. The chosen configuration of data elements is reflected to the entire LEA, including schools within the LEA. For example, a district administrator can disable the Economically Disadvantaged (ED) demographic variable from all the reports using the Administrative Configuration interface. All school and district users within the district will not be able to access the ED field on any of the reports. This configuration is specific to each district within the Commonwealth. The configuration can be changed at any time by the administrative user and the changes will be reflected instantly.

Disaster Recovery Plan

eMetric will provide "Clustered Database Services," which will enable mirroring of data on two simultaneous servers using SQL Server Clustering Services. The load-balanced web farm of application servers hosting the *Data Interaction*™ application will connect to the database cluster, thereby providing redundancy at the application and the data layers. Downtime of any single server will not cause any interruption to the service, making the downtime invisible to users. This setup requires no human intervention and provides an effective solution to mitigate major disasters. As a safeguard, a hot backup of the data warehouse will also be archived on a daily basis at an alternate location.

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

Screenshots of the Data Interaction™ System

The following are screenshots as referenced in the preceding text.

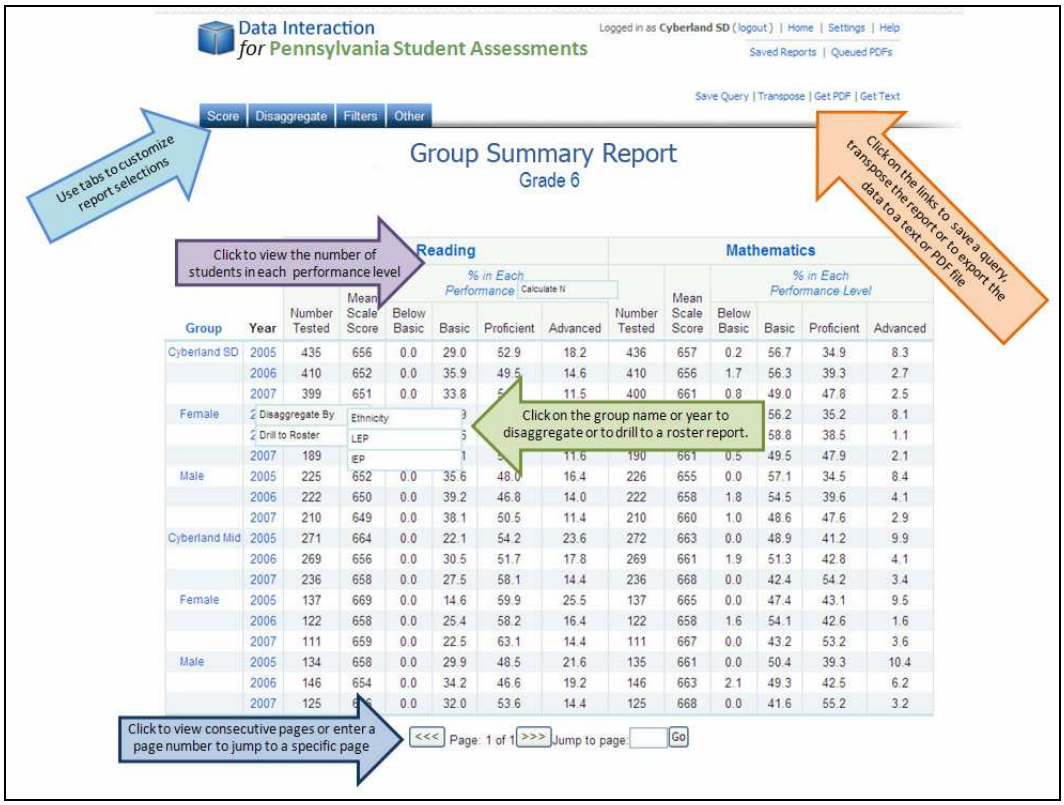


Figure 1. Group Summary Report

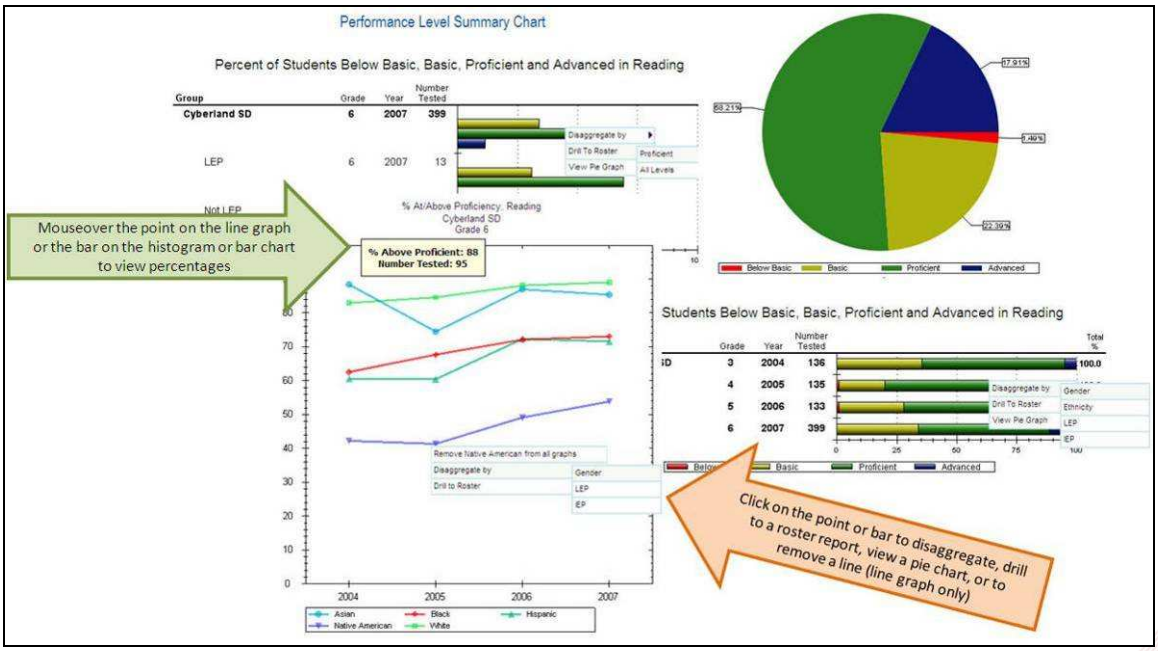


Figure 2. Graphical Summary Report

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

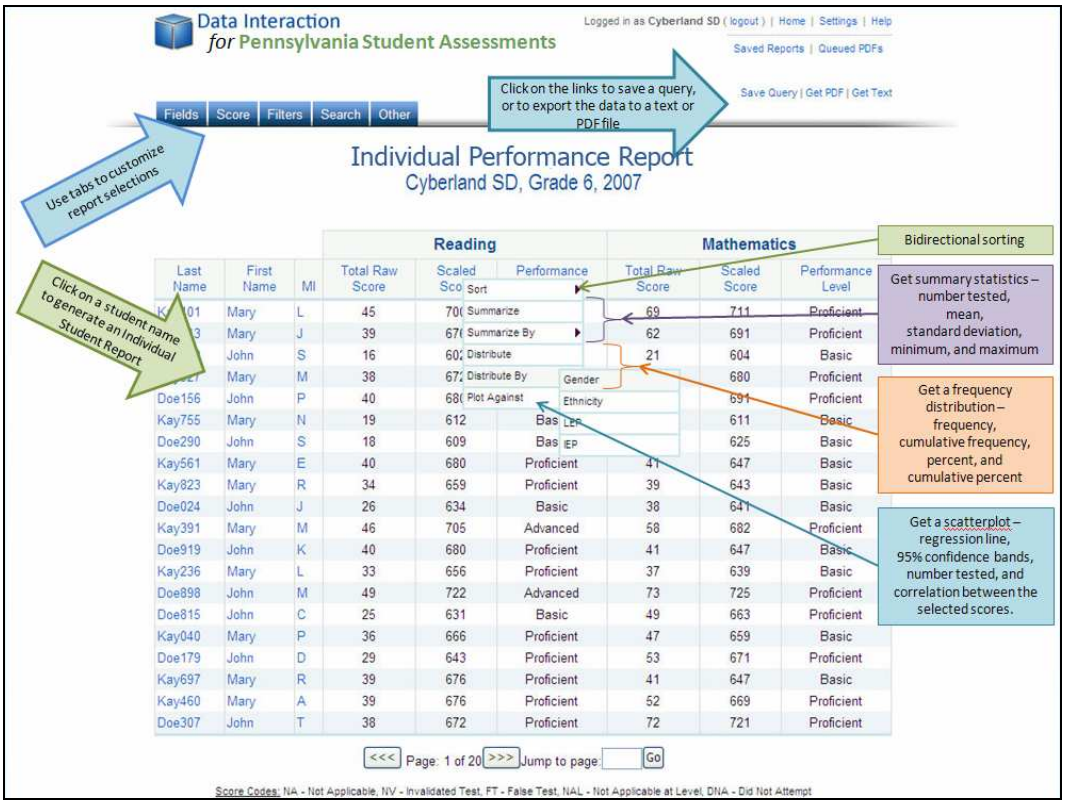


Figure 3. Individual Performance Report

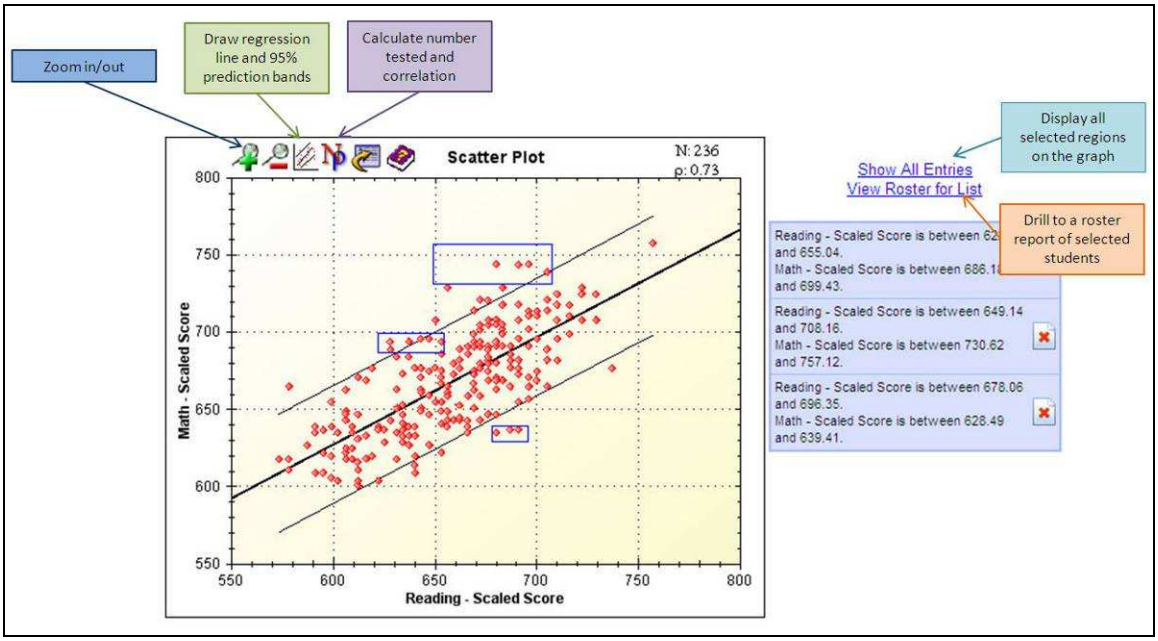


Figure 4. Scatter Plot (Plot Against feature on the Individual Performance Report)

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

Grade 5, 2006	Grade 6, 2007				
	Below Basic	Basic	Proficient	Advanced	Total
Below Basic	4	1	0	0	5
Basic	0	3	2	0	5
Proficient	0	2	74	38	114
Advanced	0	0	3	48	51
Total	4	6	79	86	175

Number Matched
Number Matched
Percent of Total
Percent of 2006
Percent of 2007

Click on any cell to drill down to a roster report of students

Figure 5. Cross – Tab Report

Longitudinal Roster Report
Cyberland Mid, Grade 6, 2004

Click on a student name to generate an Individual Longitudinal Student Report

Bidirectional sorting by latest year

Get summary statistics – number tested, mean, standard deviation, minimum, and maximum by year

Get a frequency distribution – frequency, cumulative frequency, percent, and cumulative percent by year

Sort
Summarize
Summarize By
Distribute
Distribute By
Gender
Ethnicity
LEP
IEP

Last Name	First Name	MI	Grade	Year	Reading: Total Raw Score	Reading: Scaled Score	Reading: Performance Level	Mathematics: Total Raw Score	Mathematics: Scaled Score	Mathematics: Performance Level
Doe042	John	A	7	2005	46	713	Advanced			Proficient
Doe042	John	A	8	2006	41	726	Proficient			Proficient
Doe044	John	J	6	2004	51	690	Proficient			Proficient
Doe044	John	J	7	2005	44	705	Proficient			
Doe044	John	J	8	2006	23	657	Basic	40	6	
Doe047	John	L	6	2004	45	669	Proficient	90	7	
Doe047	John	L	7	2005	46	713	Advanced	90	7	
Doe047	John	L	8	2006	41	726	Proficient	73	768	Advanced
Doe061	John	T	7	2005	14	607	Basic	32	643	Basic
Doe061	John	T	8	2006	16	633	Basic	18	640	Basic
Doe063	John	J	6	2004	28	622	Basic	36	619	Basic
Doe063	John	J	7	2005	28	652	Proficient	38	653	Basic
Doe063	John	J	8	2006	28	674	Proficient	30	669	Basic
Doe073	John	L	6	2004	49	682	Proficient	75	683	Proficient
Doe073	John	L	7	2005	40	690	Proficient	70	703	Proficient
Doe073	John	L	8	2006	38	711	Proficient	55	718	Proficient
Doe077	John	A	7	2005	26	646	Basic	33	644	Basic
Doe086	John	R	6	2004	19	597	Basic	52	645	Basic
Doe089	John	S	7	2005	45	709	Advanced	54	677	Proficient
Doe089	John	S	8	2006	33	691	Proficient	34	677	Basic

<<< Page: 2 of 38 >>> Jump to page: Go

Figure 6. Longitudinal Roster Report

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

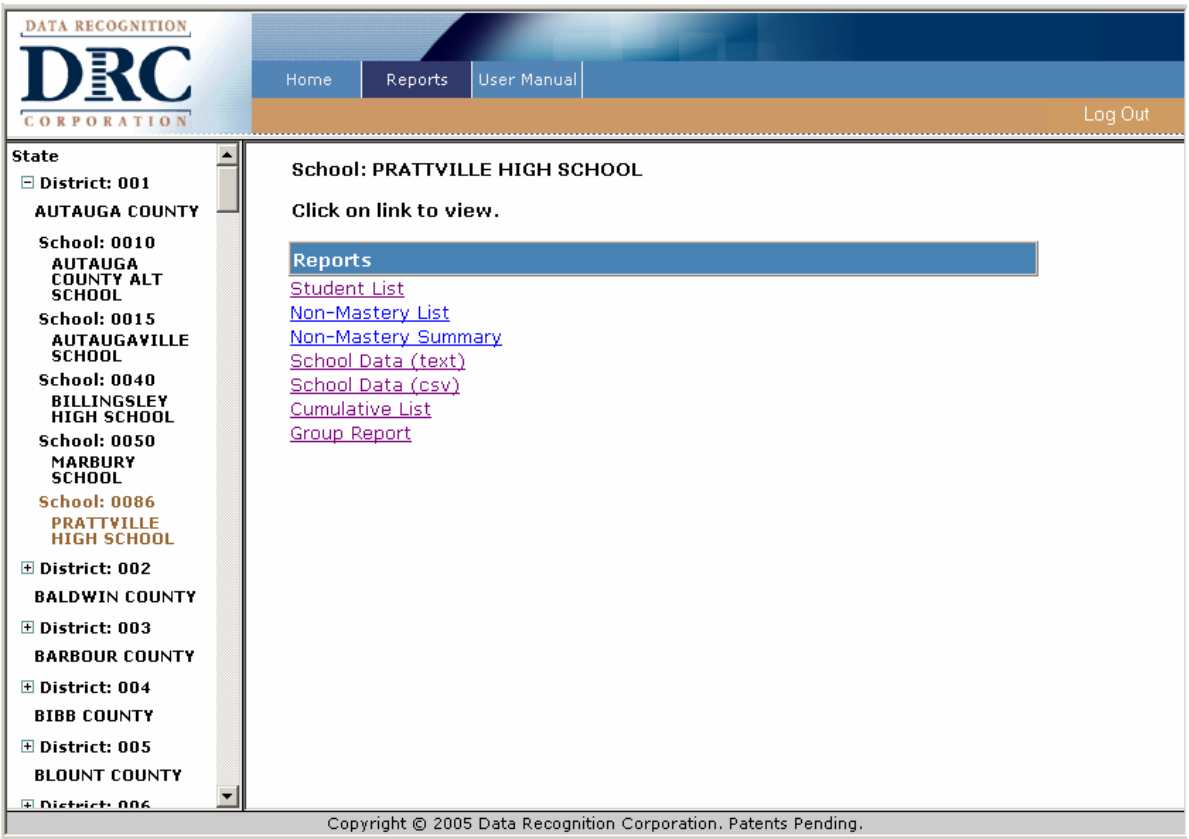
VII.K.4. Posting District Student Data Files and Summary Files

Following each administration, DRC will post district student data files on our web-based Report Delivery System for district and school users. DRC has successfully been using our Report Delivery System to post Pennsylvania assessment data files since 2004. This system provides districts and PDE the advantage of receiving data files in Excel, fixed text, or CSV format. Although the state will have access to this site, security will restrict their access to files without student identifying information.

A common file layout for the district student data file will be defined to accommodate all courses. These data files will include such information as scoring-breakdowns by reporting category and anchor, accommodations, enrollment information, and other demographic information.

As with all of our systems, **the Report Delivery System was designed with ease-of-use in mind and follows graphical user interface standards, usability guidelines, and security requirements.** District and PDE users will connect to the Report Delivery System through Ed-HUB. DRC will follow all security protocols and will ensure that our online systems are compatible with Ed-HUB.

The following figure displays a representative screen from this system.



Report Delivery System

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

To ensure the accuracy and reliability of the Report Delivery System, DRC's Software Quality Assurance Analysts, who are experienced with multiple state assessments and web-based systems, will validate that each page, link, and image displays properly. They will ensure that the system follows Graphical User Interface (GUI) standards and functions as designed.

All web system changes and modifications will be tested on a dedicated test server before being released into the production environment. The Report Delivery System will be tested on various computer platforms, using multiple browsers and numerous browser versions to ensure compatibility with the majority of the general public. Once moved to the production server, Quality Assurance Analysts will again verify that the Report Delivery System is accurate and ready for access.

VII.K.5. Sharing Data Files

DRC will be pleased to work with PDE to supply data to other vendors or educational agencies as needed under this contract. We have 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 under the GCA Program contract in the same collaborative manner.

VII.K.6. Providing Data Files to PDE

DRC has successfully defined and produced student, summary, and AYP 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.

Under the GCA Program contract, each year DRC will provide PDE with student data files. We will provide these files according to a schedule mutually agreed upon by PDE and DRC. In collaboration with PDE, DRC will define one common layout for the PDE student file that meets requirements for all subjects.

All data files will be quality checked by DRC Software Quality Assurance Analysts prior to distribution to PDE or districts.

All PDE data files will be provided to PDE for approval prior to release of data to other vendors or to the districts/schools. The printing of any documents generated from the data files will not begin until the data has been reviewed and approved by PDE.

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

VII.K.7. Disaggregated Summary Data Files

DRC will generate disaggregated summary data files as required by PDE as part of the summary level reporting process. We have been successfully providing PDE with disaggregated data files for the past five years under the PSSA contract, and DRC will provide the GCA data files using the same quality procedures that we have used in the past. For a full description of DRC's plan for summary level reporting, please see *Subheading VII.K.1.b-c*, above.

VII.K.8. Transfer of Data Files

DRC 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. DRC's Project Management, Software Development, and Quality Assurance team members will ensure that PDE has access to the required data at any point during the life of the contract.

VII.K.9. Data for Use with Item and Scoring Samplers

DRC will develop Item and Scoring Samplers each year that include released items for each 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.

VII.K.10. Access to Multiple Years of Data

DRC and eMetric's plan for providing access to multiple years of data is provided under *Subheading VII.K.3.*

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

VII.L. VALIDITY STUDIES

VII.L.1. Establishing and Documenting Validity of the GCA Scores

Validity and *reliability* have long been held to be the touchstones of educational assessment. Like most technical terms, they have a different meaning to the specialist than to the general public. To the layperson, a *reliable test* is one that can be relied on, i.e., is *worthy of trust*, which blurs the psychometric notions of reliability and validity. To make the psychometric distinction as simple as possible:

Valid means ***appropriate***;
Reliable means ***consistent***.

Test scores must be both reliable and valid before they are trustworthy.

Validity

The definition of validity currently in vogue is cited in the Standards for Educational and Psychological Testing (AERA et al., 1999, p. 9.)

Validity: *the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.*

When gathering evidence of validity, the first question to ask is "*valid for what purpose?*" In looking at validity, test developers and users must focus on exactly what **inferences** are to be drawn from the scores, and they must find evidence that supports such inferences.

A testing instrument can not, in general, be described as *valid* or *invalid*. **Test scores** may be valid for some purposes and not for others. Yardsticks and bathroom scales are both useful and consistent instruments but are not appropriate in every situation. A yardstick is not appropriate for measuring weight, and a scale is not appropriate for measuring height[1]. Neither would be *valid* for measuring artistic ability or mathematical achievement.

Traditionally, *validity* is qualified with such modifiers as *content*, *concurrent*, *predictive*, *construct*, *face*, *convergent*, *divergent*, *consequential*, and *instructional*. These modifiers do not change the main idea of validity as *appropriate for the purpose*; they do change what data are collected as **evidence** to support the validity claims. DRC's **standard practice** in our technical reports is to include evidence of content and construct validity.

[1] Actually, one might collect some evidence of *concurrent* validity to support these uses, but they usually aren't the most reliable instrument available for the purpose.

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

Ultimately, validity relies on a collection of evidence related to the quality of all technical aspects of an assessment system. The accumulation of such evidence is an ongoing process. Psychometric Services and Test Development staff at DRC have been involved in gathering validity support for tests in other states, and have developed a deep understanding of the many issues facing states in building and maintaining the validity of assessment instruments.

DRC will establish and document evidence of validity in the item development process. DRC follows these steps: (1) adherence to item specifications, (2) item writer training, (3) test editor and committee review, (4) field testing of items, and (5) post-field-test data analyses.

DRC would be happy to discuss with PDE the need for any specific, additional studies for the GCA.

Content Validity Evidence

Content validity deals with the procedures used to ensure alignment of the items with the content standards. DRC has on several occasions participated in alignment studies, using outside content specialists, to directly measure the strength of the alignment.

Evidence of content validity is the form most commonly reported with *standards-based* tests. It means the items were written to assess the specific standards and that experts agree the items align with the standards. Content validity asks the fundamental question, How well does the content of the assessment match the content of instruction?

For example, a subject-area Algebra 1 test might state that it measures the curriculum standards and core skills for the Algebra 1 course. Then every test item must show a correspondence to the curriculum standards and core skills that make up that course.

Concurrent Validity and Predictive Validity Evidence

Evidence of *concurrent* and *predictive* validity both mean that the measures from the instrument correlate well with some other measure ***of the same thing***. *Concurrent* validity typically involves correlating the test results with another measure, but this can be expensive and time consuming to collect routinely.

For example, in developing a test designed to measure computer word processing skills, which would involve speed and accuracy of key entry, the test might be given to a student who is taking a word processing course. The instructor might then be asked to observe the student's word processing skills and to rate the student using a rating sheet. The rating sheet would then be compared with the student's test results. A comparison would be made to see if the test results and the teacher ratings are in fact related. If the teacher's observational rating sheet coincides with the student's score on the test, then concurrent validity has been established.

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

Usually measures of predictive validity are obtained so that the test can be used in the future to predict success at some type of behavior—such as the ability to do the work of a key-entry word processor (McDivitt & Gibson, 2003).

Predictive validity might mean the degree to which the measures correlate with college GPA or on-the-job performance. For example, a classroom teacher might be interested in how well previous instruction predicts classroom performance next year. The teacher could examine the final exam scores from the prior year and then determine if students who scored high are getting high grades and students who scored low are getting low grades. If a correlation is found, then an inference about predicting how the students in the class will perform, based on the final exam, might be valid (McMillan, 1997). DRC has an excellent plan for collection of predictive validity evidence of the Diagnostic Assessment Tool, covered in *Subheading VIII* of this section of our proposal.

Construct and Instructional Validity Evidence

Construct validity can be gathered using the calibration and scaling information provided by the Rasch software. This information and discussion are routinely included in the technical report.

Construct validity seems to incorporate everything from the first three types (content, concurrent, and predictive) plus it relies on a solid underlying theory from psychology or cognitive science. Construct validity involves gathering evidence that there is a relationship between the content of the test and a theoretical construct it is intended to measure. Consequently, this type of validity goes beyond the simple questions of "Do the items look appropriate?" or "Do the scores predict future performance?" It asks, Do the *knowledge, skills* and *behaviors* required to do well reflect what it means to be competent in this area? It requires curriculum or instructional specialists to analyze and document the content of each item plus the thought processes needed to solve it. Evidence about construct validity also includes correlation studies to show that the measure correlates with the right things (*convergent*) and does not correlate with the wrong things (*divergent*.). Factor analysis procedures and multi-trait, multi-method matrices are often used to produce such evidence.

For many standards-based assessments tests, construct validity also involves the extent to which the test can be said to have *instructional* validity. Instructional validity is concerned with the match between what is taught in the classroom and what is actually assessed. When examining instructional validity, the major question to ask is, How closely do the test questions correspond to what has actually been taught in the classroom?

In the development of today's statewide standards-based assessment programs, the process of determining content validity, including instructional validity, often involves educators, including curriculum experts, subject-area teachers, and others. These educators, who are experts in the subject area, are asked to use their professional judgment to determine whether the test questions on a given

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

instrument do in fact measure the designated curriculum content standards or learning targets.

Instructional construct validity is a complex concept that asks several questions:

- Does the test match the standards or learning targets?
- Do the standards match the curriculum?
- Does the curriculum match the instruction?

Two important pieces of evidence for construct and instructional validity would be evidence that performance on the test:

- Improves with more instruction and student growth, and
- Changes when the instruction is changed.

Reliability

The traditional notion of test reliability involves two distinct attributes: how students in the group vary (the *standard deviation of ability*) and how precisely the test measures any one student (the *standard error of measurement.*)

Reliability: an index of how *consistently* a test orders the students.

Reliability is not concerned with whether that order is right or not. A yardstick, used in the usual way, will be equally *reliable* whether the claim is that it measures the students' height or their artistic ability.

The standard formula for an index of reliability is:

1. $$r = \frac{SD^2}{SD^2 + SE^2}$$

This index will be high (close to one) if the standard deviation of the population (*SD*) is large compared to the standard error of the instrument (*SE*). It will be close to zero if the standard deviation is small relative to the standard error. Reliability is a useful index for comparing instruments for a specific population of students.

Reliability is more relevant to *norm-referenced* tests (NRT) than *standards-based criterion-referenced* tests (CRT). NRTs focus on discriminating among students by spreading out their scores; CRTs focus on what a student can or can not do, Has this student learned this material, independent of who else was tested? Ideally, all students should be able to do everything on the CRT; consequently, their scores would not spread out at all and the test would have zero reliability

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

according to the traditional index (and no predictive or concurrent validity.) In this situation, *SD*, in equation 1, would have a value of zero.

For CRTs, the standard error of measurement (*SE*) is a more relevant indicator of a useful instrument (along with evidence of content and construct validity.) A small value of the standard error means the student's location on the construct is known very precisely. It does not matter how any other students performed.

It is axiomatic in educational assessment that there can **be no validity without reliability**. The rationale for this assertion is:

1. Reliability requires a high correlation between the test scores and whatever construct the test measures.
2. Validity requires a high correlation between the test scores and the construct that we want to measure.
3. If the test scores do not correlate well with something (i.e., unreliable), they can not correlate well with the right thing (i.e., invalid).

While this is clearly true for an NRT, it is less certain for a CRT. Since a CRT could be a very precise test (small standard error) with reliability zero, *precision* is a more appropriate precondition for validity.

DRC routinely includes traditional indices for test reliability (e.g., Cronbach's alpha) as well as standard errors of measurement for every scale score in the technical reports. DRC would be happy to discuss with PDE the need for any specific, additional studies related to reliability.

VII.L.2. Comparability Study of the Online Assessment Modality

Whenever tests that are administered under both testing modes (computer-based and traditional paper and pencil) co-exist in an assessment program, score comparability between computerized and paper-and-pencil tests becomes an important issue. Under the dual-mode testing environments, scores from the two modes cannot be used interchangeably for interpretation and/or reporting purposes without supportive evidences from carefully designed and conducted empirical research over the target testing population (AERA, APA, NCME, 1999). CAL staff has extensive experience and expertise in conducting paper-pencil (P&P) and online computer delivery (CBT) mode comparison studies within the context of high-stakes state testing environments. Descriptions of various comparability designs (i.e., double testing or test-retest, matched groups, volunteer groups, and randomly assigned groups) are presented below, followed by the proposed comparability study for the GCA administrations.

CAL staff has conducted paper-pencil (P&P) and online computer delivery (CBT) mode comparison studies yearly over the life of the online testing program in the state of Kansas. The design and results from the studies during the first year (2003) for the grade 7 mathematics test can be found in Poggio, Glasnapp, Yang

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

and Poggio (2005) at http://www.jtla.org. The design and results from the studies during the second year (2004) for tests at three grade levels of both reading and mathematics tests can be found in Poggio, Glasnapp, Yang, Beauchamp and Dunham (2005). When these studies were conducted, the only viable design for data collection was to "double test" students on parallel forms of the test, once under a P&P format and once under the CBT format. District/school participation in the studies was voluntary. Although results lacked statistical difference based on very large sample sizes (numbering in the tens of thousands), a limitation of these two studies was that order of testing mode was not controlled nor was the order data reported by schools viewed to be trustworthy.

The double tested design and data collection were replicated again during the 2005 testing period for tests at three grade levels in reading and mathematics, but data also were collected for test forms at three grade levels in science and social studies. Information to study and control the order effect was captured in this series of studies. In this work, again we did not observe statistically significant results between the computerized testing and paper and pencil modes for grades or content areas.

In the Kansas comparability studies, the conditions and constraints of the testing program when the studies were initiated necessitated that a "double testing" design be put in place so that data would exist such that the individual students served as their own control in the repeated measures design. In this design, order of administration is a potential problem and would best be controlled through random assignment. Because of the administration time involved in double testing and the uncertainty of having adequate controls for the order effect through random assignment of order, more efficient data collection designs are available if the right conditions exist in a state's testing program.

An alternative and attractive quasi-experimental design does exist if the right conditions can be put in place. If students' prior years' test scores are available, these scores can be used as matching control variables or covariates to control for potential prior achievement differences in the volunteer CBT group and the selected P&P comparison group along with other matching covariates (propensity scores for matching and controlling). While the immediate prior year test scores in Pennsylvania would not be available for operational GCA assessments, prior scores from earlier Diagnostic Assessments likely would be available and could be used as matching control variables along with select demographic variables to control for achievement differences in the groups taking tests under the different mode conditions. This approach would have considerable merit and value during early years of operational implementation.

Another study possibility is to use the data as it exists from volunteer groups knowing that they likely represent non-equivalent groups, but attempt to demonstrate comparability by looking at the structural consistency of the tests across modes, (i.e., conduct the studies addressing structural validity with test mode as a variable). If differences are not found, evidence is provided to support a conclusion that the tests are functioning similarly. Such approaches have been

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

taken in the Kansas studies (2006) and have demonstrated an exceptionally high degree of structural consistency of the test items across modes.

There is no doubt that the best, failsafe design is to implement a true randomized experimental design with random selection and assignment of student to test mode. It is the design of choice and the one we would recommend is implemented if at all possible. However, random assignment of administration mode (paper and pencil or online) is preferable at the student level, but often is not practical. If this design is desired by PDE, we will work with the Department to design the most feasible random sampling and assignment design (at the student, class, or building level) to be implemented during test administrations. The assigned groups would be of sufficient size and representativeness as to be considered randomly equivalent. We would then explore the structural similarity of the constructs being assessed by the GCA assessments across delivery modes through the use of appropriate factoring techniques, similar to the procedures employed in the volunteer groups design described above.

In addition to the work by CAL personnel, it should also be noted that DRC was awarded a contract from the state of South Carolina (2006-07) to explore the feasibility of moving P&P testing to the online CBT delivery mode. As part of that contract, an exhaustive literature review was made examining studies addressing the comparability issue. This latter review will serve as a foundation to provide CAL, DRC, and PDE with information in deciding which design to best implement within the context of the Pennsylvania GCA online assessment implementation schedule.

This presentation on comparability is intended to convey the vast experience and expertise of CAL and DRC staff in addressing the comparability issue. DRC and CAL will work with PDE to implement the most efficient and valid design within the context of the Pennsylvania GCA testing program during the contract period to address the comparability of paper and pencil and online test delivery mode. In the end, we would plan to implement a design that PDE and its advisors wholeheartedly support and endorse.

VII.M. Assessment Schedule

VII.M.1. Schedule of Proposed Assessment Windows and Providing Assessment Results

DRC is pleased to present, for PDE's consideration, our proposed GCA assessment window and reporting recommendations for 2011, 2012, and 2013. DRC has met or exceeded PDE's performance requirements, including:

- Reporting of results for graduating seniors no less than 10 calendar days prior to graduation.
- Reporting of results for all other students before the end of the school year.
- School and district results as quickly as possible.

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

Assessment Windows

DRC is proposing that all assessments be given in one-week time periods: the first week of May for the Spring administration; in early August for the Summer administration; and in early December for the Fall administration. Each course will be assigned a specific two-day testing window within the designated week-long GCA window.

Materials Delivery and Return

DRC is proposing that administrative materials for all GCA courses arrive in the schools and districts four weeks prior to the beginning of the testing window. Similarly, we are proposing that all secure materials for all courses be shipped together in one shipment to arrive two weeks prior to the testing window. DRC is recommending this approach for materials shipping and delivery to ease the handling burden for schools and districts and to keep costs down. Other alternatives can be discussed and negotiated upon award.

Materials will be returned to DRC by course after each two-day window. To ease the management burden on school and district personnel, DRC is proposing that materials be shipped directly from all schools via UPS. We have found this to be the easiest method of return shipping for school personnel, since many schools have daily UPS pick-ups. Additional discussions of our proposed collection processes can be found under *Subheading VII.H., Assessment Materials Collections and Shipping*.

Score Reporting

DRC understands PDE's desire to have results in the field in an expeditious manner. We commend PDE for challenging offerors to devise a reporting plan to accomplish this goal. DRC is pleased to present our proposed model for reporting in the timeframes required by the RFP. For additional detail regarding our proposed reports, please see *Subheading VII.K., Reports and Data Files.*

DRC proposes the following reporting model for all subjects:

- Report scores (based on scores for both multiple-choice and open-ended items) for graduating seniors provided no less than 10 calendar days prior to graduation.
- Report scores (based on scores for both multiple-choice and open-ended items) for all non-graduating students provided before the end of the school year.
- Printed ISRs will reflect both multiple-choice and open-ended scores and will be delivered in July.

Please refer to *Subheading VIII.K., Reports and Data Files,* for more information about our proposed approach for reporting based on multiple-choice items.

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

Unique Considerations for Reporting Scores for First-Time Administrations in December

For the first operational administrations that occur before standard setting (December 2010 and Spring 2011 for Wave 1 courses and December 2011 and Spring 2012 for Wave 2 courses), DRC proposes the following model:

- Provide a preliminary raw-score-only roster report via DRC's Report Delivery website.
- Provide full scores (scale scores and performance levels) after standard setting is complete, following the Spring administration.

DRC recommends that standard setting be conducted after the Spring administration because the population of test takers will be higher and, therefore, statistics generated from the assessment will be more reliable for the purpose of setting standards. Please refer to *Subheading VII.J.8., Conducting Standard Setting Sessions*, for additional information about standard setting.

Timeline Charts

The charts below outline DRC's proposed testing windows and reporting dates for the 2011, 2012, and 2013 Spring assessments.

Some important interim dates have been included in the charts (e.g., processing begins, handscoring ends), to help illustrate the entire process—from testing to reporting. Testing windows and reporting dates for the Summer and Fall administrations would follow the same timelines. DRC understands the size and complexity of the GCA program. The management of the processing, scoring, and reporting, with critical turnarounds, is no small feat. It takes tremendous effort to orchestrate all the required tasks—DRC understands this and believes we are in the best position to accomplish this for years to come.

DRC would be happy to discuss alternative windows and scheduling with PDE upon contract award.

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

DRC's Proposed Three-Year Testing Window and Reporting Schedules

2011 Schedule First Year Algebra I/II, Geometry Operational		
Description	Non-12th Graders*	12th Graders*
Materials Delivery		
Administration materials arrive at districts/schools	By April 4	By April 4
Secure materials arrive at districts/schools	By April 18	By April 18
Testing Windows		
Geometry	May 2–3	May 2–3
Algebra I	May 3–4	May 3–4
Algebra II	May 4–5	May 4–5
Make-Up Tests	May 6–9	May 6–9
Materials Pick-Up From Schools		
Geometry	May 4	May 4
Algebra I	May 5	May 5
Algebra II	May 6	May 6
Make-Up Tests	May 10	May 10
Materials Processing		
Geometry Processing	May 5–24	May 5–16
Algebra I Processing	May 6–24	May 6–16
Algebra II Processing	May 9–24	May 9–16
Handscoring (Operational OE Items Only)		
Geometry OE Scoring	May 6–27	May 6–19
Algebra I OE Scoring	May 9–27	May 9–19
Algebra II OE Scoring	May 10–27	May 10–19
Raw Score Reporting (MC and OE)		
Algebra I, Algebra II, and Geometry raw score test results available in field	**May 31**	**May 23**
Standard Setting and State Board Approvals		
Algebra I, Algebra II, Geometry Standard Setting	June 7–10	June 7–10
State Board approval of Algebra I, Algebra II, and Geometry scaled score cut points	June 16	June 16
Student Reporting (MC and OE)		
Algebra I, Algebra II, and Geometry MC and OE results to PDE	June 22	June 22
PDE approves Algebra I, Algebra II, and Geometry MC and OE results	June 23	June 23
Algebra I, Algebra II, and Geometry results available in field	**June 24**	**June 24**
Individual Student Reports arrive in districts (hard-copy)	**July 22**	**July 22**

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

2011 Schedule First Year Algebra I/II, Geometry Operational		
Description	Non-12th Graders*	12th Graders*
Summary Reporting (Online)		
PDE receives summary data files and final School and District Summary Reports	July 1	July 1
PDE approves summary data files and final School and District Summary Reports	July 8	July 8
School and District Summary Reports available in field	**July 11**	**July 11**
Data Interaction™		
PDE receives *Data Interaction*™ tool	July 1	July 1
PDE approves *Data Interaction*™ tool	July 8	July 8
***Data Interaction*™ tool available in field**	**July 11**	**July 11**
Missing Materials Reporting		
Missing Materials Report provided to PDE	May 30	May 30
Technical Reporting		
Technical Reports delivered to PDE	September 30	September 30
*Note that for the purposes of these schedules, DRC has assumed a June 1 graduation and end-of-school year date.		

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

2012 Schedule Second Year Algebra I/II, Geometry Operational First Year Biology, World/US History, English Comp., Literature, Chemistry, Civics & Government Operational		
Description	Non-12th Graders*	12th Graders*
Materials Delivery		
Administration materials arrive at districts/schools	By April 9	By April 9
Secure materials arrive at districts/schools	By April 23	By April 23
Testing Windows		
English Composition and Literature	April 30–May 2	April 30–May 2
Algebra I, Algebra II, Geometry	May 1–2	May 1–2
US History, World History, and Civics & Government	May 2–3	May 2–3
Biology and Chemistry	May 3–4	May 3–4
Make-Up Tests	May 7–9	May 7–9
Materials Pick-Up From Schools		
English Composition and Literature	May 3	May 3
Algebra I, Algebra II, Geometry	May 3	May 3
US History, World History, and Civics & Government	May 4	May 4
Biology and Chemistry	May 7	May 7
Make-Up Tests	May 10	May 10
Materials Processing		
English Composition and Literature	May 4–22	May 4–15
Algebra I, Algebra II, Geometry	May 4–22	May 4–15
US History, World History, and Civics & Government	May 7–23	May 7–16
Biology and Chemistry	May 8–23	May 8–16
Handscoring (Operational OE Items Only)		
English Composition and Literature	May 8–25	May 8–18
Algebra I, Algebra II, Geometry	May 9–25	May 9–18
US History, World History, and Civics & Government	May 10–25	May 10–18
Biology and Chemistry	May 11–25	May 11–18
Raw Score Reporting (MC and OE)		
Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government raw score test results available in field	**June 1**	**May 22**
Standard Setting and State Board Approvals		
Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government Standard Setting	June 5–8	June 5–8

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

2012 Schedule Second Year Algebra I/II, Geometry Operational First Year Biology, World/US History, English Comp., Literature, Chemistry, Civics & Government Operational		
Description	Non-12th Graders*	12th Graders*
State Board approval of Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government scaled score cut points	June 14	June 14
Student Reporting (MC and OE)		
Algebra I, Algebra II, Geometry MC and OE results to PDE	May 30	May 22–AM
PDE approves Algebra I, Algebra II, Geometry MC and OE results	May 31	May 22–PM
Algebra I, Algebra II, Geometry results available in field	**June 1**	**May 22–PM**
Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government MC and OE results to PDE	June 19	June 19
PDE approves Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government MC and OE results	June 21	June 21
Biology, Chemistry, English Composition, Literature, US History, World History, and Civics & Government results available in field	**June 22**	**June 22**
Individual Student Reports arrive in districts (hard–copy)	**July 20**	**July 20**
Summary Reporting (Online)		
PDE receives summary data files and final School and District Summary Reports	June 29	June 29
PDE approves summary data files and final School and District Summary Reports	July 6	July 6
School and District Summary Reports available in field	**July 9**	**July 9**
Data Interaction™		
PDE receives *Data Interaction*™ tool	June 29	June 29
PDE approves *Data Interaction*™ tool	July 6	July 6
***Data Interaction*™ tool available in field**	**July 9**	**July 9**
Missing Materials Reporting		
Missing Materials Report provided to PDE	May 30	May 30
Technical Reporting		
Technical Reports delivered to PDE	September 28	September 28
*Note that for the purposes of these schedules, DRC has assumed a June 1 graduation and end-of-school year date.		

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

2013 Schedule All courses operational with no standard setting.		
Description	Non-12th Graders*	12th Graders*
Materials Delivery		
Administration materials arrive at districts/schools	By April 1	By April 1
Secure materials arrive at districts/schools	By April 15	By April 15
Testing Windows		
English Composition and Literature	April 29–May 1	April 29–30
Algebra I, Algebra II, Geometry	April 30–May 1	April 30–May 1
US History, World History, and Civics and Government	May 1–2	May 1–2
Biology and Chemistry	May 2–3	May 2–3
Make-Up Tests	May 6–8	May 6–8
Materials Pick–Up From Schools		
English Composition and Literature	May 2	May 2
Algebra I, Algebra II, and Geometry	May 2	May 2
US History, World History, and Civics and Government	May 3	May 3
Biology and Chemistry	May 6	May 6
Make-Up Tests	May 9	May 9
Materials Processing		
English Composition and Literature Processing	May 3–22	May 3–15
Algebra I, Algebra II, and Geometry Processing	May 3–22	May 3–15
US History, World History, and Civics and Government Processing	May 6–22	May 6–15
Biology and Chemistry Processing	May 7–22	May 7–15
Handscoring (Operational OE Items Only)		
English Composition and Literature OE Scoring	May 7–24	May 7–16
Algebra I, Algebra II, and Geometry OE Scoring	May 8–24	May 8–16
US History, World History, and Civics and Government OE Scoring	May 9–24	May 9–16
Biology and Chemistry OE Scoring	May 10–24	May 9-16
Student Reporting (MC and OE)		
Algebra I, Algebra II, Biology, Chemistry, Civics and Government , English Composition, Geometry, Literature, US History, and World History MC and OE results to PDE	May 29	May 22–AM
PDE approves Algebra I, Algebra II, Biology, Chemistry, Civics and Government , English Composition, Geometry, Literature, US History, and World History MC and OE results	May 30	May 22–PM
Algebra I, Algebra II, Biology, Chemistry, Civics and Government , English Composition, Geometry, Literature, US History, and World History results available in field	**May 31**	**May 22–PM**

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

2013 Schedule All courses operational with no standard setting.		
Description	Non-12th Graders*	12th Graders*
Individual Student Reports arrive in districts (hard-copy)	**July 5**	**July 5**
Summary Reporting (Online)		
PDE receives summary data files and final School and District Summary Reports	June 10	June 10
PDE approves summary data files and final School and District Summary Reports	June 14	June 14
School and District Summary Reports available in field	**June 17**	**June 17**
Data Interaction™		
PDE receives *Data Interaction*™ tool	June 10	June 10
PDE approves *Data Interaction*™ tool	June 14	June 14
***Data Interaction*™ tool available in field**	**June 17**	**June 17**
Missing Materials Reporting		
Missing Materials Report provided to PDE	June 5	June 5
Technical Reporting		
Technical Reports delivered to PDE	September 30	September 30

*Note that for the purposes of these schedules, DRC has assumed a June 1 graduation and end-of-school year date.

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

Works Cited

Allen, N. L., Carlson J. E., & Zelenak, C. A., (1999). *The NAEP 1996 technical report* (NCES 1999 – 452). Washington, DC: National Center for Education Statistics.

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.

Angoff, W. H. (1971). Scales, norms and equivalent scores, in R.L. Thorndike (Ed.), *Educational measurement*, (2nd edition, pp. 508–600), Washington DC: American Council on Education.

Angoff, W. H. (1974). The development of statistical indices for detecting cheaters. *Journal of the American Statistical Association, 69,* 44–49.

Birch, M. W. (1964). The detection of partial association I: The 2x2 case. *Journal of the Royal Statistical Society, B26,* 313–324.

Cook, L. & Eignor, D. (1991). IRT equating methods. *Educational Measurement: Issues and Practice. 10,* 37–45.

Fletcher, R. B. (2000). *A review of linear programming and its application to the assessment tools for teaching and learning projects.* (Technical Report 5, Project as TTle). Auckland, New Zealand: University of Auckland.

Hanson, B. A. & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27,* 345–359.

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenzsel procedure. In H. Wainer and H. I. Braun (Eds.), *Text Validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kingston N. M., Kahl, S. R., Sweeny, K. P., & Bay L., (2001). Setting Performance standards using the Body of Work method. In G. J. Cizek (Ed.), *Setting Performance Standards* (pp. 219–148). Mahwah, NJ: Lawrence Erlbaum Associates.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*: *Methods and Practices*. (2nd ed). New York: Springer–Verlag.

Lewis, D. M., Mitzel, H. C., Green, D. R. & Patz, R. J. (1999). *The Bookmark standard setting procedure*. Monterey, CA: CTB/McGraw–Hill.

Linacre, J. M. (2008). WINSTEPS. [Computer Program]. Chicago: WINSTEPS.com.

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

Linn, R. L. (1993). Linking the results of distinct assessments. *Applied Measurement in Education, 6(1),* 83 –102.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32(2),* 179–197.

Lord, M. (1977). Practical application of item characteristic curve theory. *Journal of Educational Measurement, 14(2),* 117–138.

Ludlow, L. H. (1986). Graphical analysis of item response theory residuals. *Applied Psychological Measurement, 10*, 217–229.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22,* 719 –748.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika 47*, 149–174.

McDivitt, P.J. & Gibson, D. (2203). Guidelines for selecting appropriate tests. In Wall, J.E. and Walz, G. H. (eds.), *Measuring Up: Assessment issues for teachers and counselors, and administrators*. Greensboro, NC: Educational Resources Information Center Counseling and Student Services Clearinghouse and the National Board for Certified Counselors.

McMillan, J. H. (1997). *Classroom assessment; principles and practice for effective instruction*. Needham Heights, MA: Allyn & Bacon.

Mead, R. J. (1976). *Assessing the fit of data to the Rasch model through the analysis of residuals*. Unpublished doctoral dissertation, Chicago: University of Chicago.

Mead, R. (1980). Using the Rasch model to identify person–based measurement disturbances. *Proceedings of the 1979 Computer Adaptive Testing Conference*. Minneapolis, MN: University of Minnesota.

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

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Smith, R. M. (1982). *Detecting measurement disturbance with the Rasch model*. Unpublished doctoral dissertation. University of Chicago

Smith, R. M. (1986). Person fit in Rasch model. *Educational and Psychological Measurement, 46*, 359–372.

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

Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, *1*, 199–218.

Smith, E. V. & Smith, R. M. (2004). *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.

Smith, E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation, in E. V. Smith & R. M. Smith (Eds.) *Introduction to Rasch measurement*, (pp. 93–122). Maple Grove, MN: JAM Press.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved March 15, 2008 from: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 278–286.

Webb, N. L. (2002). Alignment study in language arts, mathematics, science, and social studies of state standards and assessment for four states. Washington, DC: Council of Chief State School Officers.

Wright, B. D. (1968). Sample–free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.

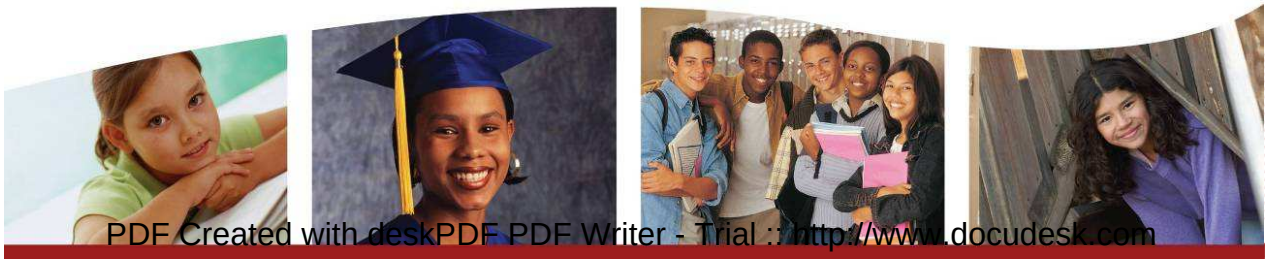
Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D. & Stone, M. (1979). *Best test design*. Chicago: MESA Press.

Zwick, R., & Thayer, D. T. (1996). Evaluating the magnitude of differential item functioning in polytomous items. *Journal of Educational and Behavioral Statistics 21*,187– 201.

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

DATA RECOGNITION
DRC
CORPORATION



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