TR 2008-16

Ohio Achievement Test

Operational Technical Report for Grade 3 Reading, October 2008 Adminstration

American Institutes for Research

December 31 2008

Revision	Description	Revised by	Approval	Date
Int. Rel.	Edited by T. Moore	TWM	DX	12.31.08

TR 2008 – 16

OHIO ACHIEVEMENT TEST

GRADE 3 READING

OCTOBER 2008 ADMINISTRATION

OPERATIONAL TECHNICAL REPORT

AMERICAN INSTITUTES FOR RESEARCH

NOVEMBER 7, 2008

Table of Contents

Appendices

Background

The purpose of this report is to summarize the results of statistical and psychometric analyses performed on the operational data from the October 2008 administration of the Ohio Achievement Tests for grade 3 reading. Before summarizing the results of the October 2008 operational assessment, we begin with an overview of the processes involved in designing and administering the Ohio Achievement Tests to provide the necessary context for a more complete understanding of the test results.

Ohio Achievement Tests

The Ohio Achievement Tests represent a system of standards-based achievement tests administered in grades 3 through 8 in five subject areas: reading, math, science, social studies, and writing. The reading and mathematics assessments are administered annually in grades 3 through 8. The science and social studies achievement tests are administered to all Ohio public school students in grades 5 and 8, and the writing achievement tests are administered to students in grades 4 and 7. In addition, the grade 3 reading test is administered twice a year, once at the beginning of the academic year, and a second time in the spring near the end of the academic year. Students scoring at the proficient level or higher in the fall administration of the grade 3 reading test are not required to participate in the spring administration of the grade 3 reading test. Previously, students failing to meet the proficient standard in either the fall or spring administration were required to participate in a summer administration of the grade 3 reading test, but this test is no longer being offered by ODE.

Ohio Academic Content Standards Development

The Ohio Academic Content Standards constitute the basis of the Ohio state assessment program; they indicate what students should know and be able to do for every grade and subject area. In 1997, the Ohio Department of Education, in conjunction with the Ohio Board of Regents, convened teams of teachers, parents, and community members for each of the content areas to work toward creating a set of common expectations for each subject. These teams drafted statements that clearly defined learning expectations for each grade level. Prior to state-wide implementation into school curricula, these academic content standards were subjected to an extensive review process. Following an initial draft of the standards, professional organizations with content specialties in the different subject areas had the opportunity to review the standards. Following this phase, the draft academic content standards were reviewed through a series of public engagement activities designed to elicit feedback from stakeholders. The draft academic content standards were then released for public review. The final step consisted of review and adoption of the standards by the Ohio State Board of Education, and integrating them into the state-wide curricula.

The Ohio Academic Content Standards are supported by benchmarks and grade-level indicators, which clarify the standards and provide more specific information regarding the content for which the students are responsible. The reader is directed to the Ohio Department of Education (ODE) Web site (http://www.ode.state.oh.us/), which houses the Ohio Academic Content Standards for all grades and subject areas.

Item and Test Specification Development

Following the adoption and integration of the Ohio Academic Content Standards into the school curricula, item and test specifications were developed to make sure that the tests and their items are aligned to the standards, benchmarks, and grade-level indicators they are intended to measure. These item and test specifications identify the item types, quantity, and point values to be included in the assessments. These specifications also include the distribution of items across content standards, including the number of items and score points required for the measurement of each content standard and benchmark. Specifications for reading tests include rules for identifying and selecting appropriate reading passages. Test specifications also designate test characteristics such as test and item complexity, in order to guarantee that all achievement tests include items of varying degrees of difficulty. The ODE Web site (http://www.ode.state.oh.us/) provides access to test blueprints.

Performance Level Descriptors (PLDs) were developed to aid educators and test developers in understanding the nature of how the academic standards are manifested in student performance. PLDs define the content area knowledge, skills and processes that examinees at a performance level are expected to possess. ODE's descriptions of Limited, Basic, Proficient, Accelerated and Advanced performance are public statements about what and how much Ohio educators want students to know and be able to do by the end of each grade and in each subject area. Thus, the PLDs are intended to provide a clear link between the test content and the Ohio Academic Content Standards and their corresponding performance levels. PLDs for each of the Ohio Achievement Tests are available at the ODE website (http://www.ode.state.oh.us/).

Item Development

The next required step is to develop items that measure the academic standards. All items pass through an extensive review process, including a series of internal AIR reviews, internal ODE reviews and Ohio Fairness and Content Advisory Committee reviews, before inclusion first on field test forms, and subsequently on operational test forms. Content specialists at AIR initially write the items, which then must pass though several internal review stages, including content, editorial, and senior content reviews. Items that are reviewed and approved internally are then sent to ODE for their review. ODE reviews the items and provides an outcome (Accept as Appears, Accept as Revised, Revise and Resubmit, or Reject) for each item. AIR and ODE then discuss the requested revisions and ODE signs off on each item as ready for Ohio Committee review.

Following the completion of the AIR and ODE internal item development cycle, ODE then convenes two committees, the Fairness and Sensitivity Committee, which consists of Ohio teachers and community members, and the Content Advisory Committee, which is comprised of Ohio teachers from across the state. The Fairness and Sensitivity Committee review ensures that the items remain free from bias or stereotype, while the Content Advisory Committee review determines whether the items are properly aligned to the content standards, benchmarks, and grade-level indicators; accurately measure intended content; and are grade level appropriate. Following approval from both committees, the items are then added to the field test pool so that they can be used on a future field test form.

Field-testing. To construct field test forms, AIR content experts work with ODE curriculum and assessment experts to select items from the field test pool to meet the requirements described in the test blueprints. AIR uses two different form designs to conduct field tests of new items within the Ohio assessment system, the *common item block design* and the *embedded item design*.

For independent field tests, where many items are being field-tested to establish an operational item bank, we employ a *common item block design*. The common item block design maximizes the number of items field-tested—while minimizing the exposure of field-test items in the event that a field-test form(s) is compromised. This design allows for common item equating while isolating common item blocks so that they only appear on a limited number of forms. This minimizes the number of items that could be compromised in case a field test form is breached.

The second form design is an *embedded item design* that is used for grades and subjects in which operational tests currently exist. Once operational forms have been established, it is necessary to continue to replenish the operational item pool in order to adjust for items that have been breached or released to the public. In this design, a small subset of items is field tested as part of the operational form.

The sampling procedure for the field-test forms results in a random sample of students taking each of the achievement tests. For independent field tests, schools are selected within each stratum based on a simple random sample scheme. All classes are selected from each sampled school, and all students within each selected classroom are selected. The series of field test technical reports provides a more detailed look at the sampling procedures, including the sampling strata used in the Ohio assessment system and the steps involved with selecting schools for the field test samples.

Item analyses and data review. Following each field test administration, classical and item response theory (IRT) statistical analyses are performed on student response data. The item analyses provide information about the quality of the items. Items are flagged for review for the following reasons:

- Proportion correct value is less than .25 or greater than .95 for multiple-choice items, or greater than .95 for any single score point of a constructed-response item;
- Adjusted biserial/polyserial correlation statistic is less than .25 for multiple-choice or constructed-response items;
- Adjusted biserial correlations for multiple-choice item distractors is greater than .05;
- The proportion of students responding to a distractor exceeds the proportion responding to the keyed response for MC items;
- Mean total score for a lower score point exceeds the mean total score for a higher score point for constructed-response items;
- Omit rate is greater than .15; and
- The item falls into the C category for any differential item functioning (DIF) contrast. The C category indicates evidence of significant DIF and is defined, for dichotomous

items, as $MH\chi^2$ is significant and $|\hat{\Delta}_{MH}| \geq 1.5$, and for polytomous items $MH\chi^2$ is significant and $|SMD|/|SD| \geq .25$.

AIR also conducts DIF analyses on all items included in the field test to detect potential item bias across major ethnic and gender groups. The performance on each item by subgroup members (Black/African American students, Hispanic students, Multi-Ethnic students, and female students) is compared with the performance of the appropriate reference group (White students or male students), resulting in four sets of comparisons: Black/White, Hispanic/White, Multi-Ethnic/White, and female/male. The purpose of these analyses is to identify items that may favor students in one group over those of similar ability in another.

Items flagged for review on the basis of any of the aforementioned criteria must pass a three-stage data review process to be included in the final item pool from which operational forms are created. As a first level of review, a team of AIR psychometricians reviews all flagged items to ensure that the data are accurate, properly analyzed, have correct response keys, and have no obvious problems with the items.

Second, ODE curriculum and assessment specialists review the item statistics and content appropriateness of all flagged items. Additionally, a Fairness and Sensitivity Committee reviews all items flagged on the basis of DIF statistics. Committee members are encouraged to discuss these items with the statistics as a guide, and are asked to make decisions regarding whether the item should be excluded from the pool of potential items given its performance in field testing.

Operational Forms Construction

As with test items, operational forms pass through a programmed sequence of review levels before they can be administered. Test development specialists select items from the operational pool to match the test specifications.

After test content specialists have developed an operational form, it must be submitted for review by AIR psychometricians. Psychometricians evaluate each form to determine whether the test form meets specified statistical criteria. Test characteristic curves are evaluated to ensure that test characteristic curve differences meet tolerances to base year operational forms so that test information across the range of test scores is similar across test administrations. Tolerances for form difficulties are also evaluated to ensure that raw scores at the proficiency cut score remain consistent across test administrations. Checks are also performed to make sure that forms meet test specifications for the number of items and points, both overall and by content standard.

After receiving psychometric approval for a proposed operational test form, the form is submitted to a senior test development specialist to ensure that test specifications for distribution of item content and item type are met. Following senior test development specialist review and approval, proposed operational forms are submitted to ODE for review. Any revisions to the form, whether during senior test development or ODE review, require that the form be resubmitted for evaluation by AIR psychometricians. Following final ODE approval of an operational form, the form is available for use in an operational assessment.

Standard Setting

Performance standards for the Ohio Achievement Tests were recommended through a series of standard-setting workshops conducted as each test in the system became operational. Detailed descriptions of the procedures and results for each standard setting workshop are provided in a series of standard setting technical reports available from ODE. Each of the standard setting workshops generally followed the same set of procedures reviewed here.

In each case, the goals of the standard-setting workshop panelists were to:

- make initial placements of recommended cut scores on the Ohio Achievement Tests corresponding to the Performance Level Descriptors for Basic, Proficient, Accelerated and Advanced levels of performance;
- consider agreement and impact data to guide judgments about item difficulty and placement of the bookmarks; and
- make final recommendations to ODE about the appropriate placement of Basic, Proficient, Accelerated and Advanced performance levels for each of the tests.

Panel composition. With ODE's direction, assistance, and approval, AIR recruited a set of panelists for each of the standard-setting workshops. The panelists' background characteristics were intended to be quite similar to those of the panelists who participated in earlier standard-setting workshops for Ohio assessments.

In recruiting panelists, AIR sought representation of females and males as found in the teacher population in Ohio. The same principle was applied to the geographical representation of panelists, with members recruited from the northwest, northeast, central southeast, and southwest sections of Ohio. In addition, we sought proportional representation of African American, Hispanic, Asian, and white members; members from urban, suburban and rural school systems; and members from school systems in high-, moderate- and low-income communities.

Within each of the subject area and grade level panels, participants were initially assigned to one of four or five tables, with each table seating five panelists representing teachers, other educators and community representatives, according to the recruitment design described above. Following the large group training on the first day of standard setting, some panels needed to consolidate panelists into four tables to account for absentee panelists.

AIR worked with ODE staff to identify candidate Table Leaders prior to the workshop. AIR convened all Table Leaders prior to the start of the workshops to train the Table Leaders in the Bookmark method and explain their roles and responsibilities. In general, the Table Leader's role was to work with the standard setting staff to facilitate discussion, share insights, provide information to the panelists, report any concerns, collect all data sheets and secure materials, and ensure that panelists carry out their roles effectively.

Training. Training consisted of a review and discussion of the Ohio Academic Content Standards, the test blueprints, and the PLDs for each performance standard. A general overview of the standard setting workshop and the Bookmark method was provided to all panelists as part of large group training. Participants reviewed the scoring procedures, scaling procedures and

other details of the testing process that were necessary for recommending performance standards. They learned about the role of response probabilities in placing their bookmarks and how to apply the appropriate RP criterion for placing bookmarks in their Ordered-Item Booklets.

Panelists internalized the concept of students who are "just barely Proficient" (and Basic, Accelerated and Advanced) and the bookmark placement task: place the bookmark on the page where you would expect one-half of students who are just barely Proficient (and Basic, Accelerated and Advanced) to respond successfully. The training was organized into two parts: a) a general overview of standard setting and training on the Bookmark procedure and b) a specific orientation to the Ohio content standards, test items, scoring criteria and PLDs. The session began with a review of the purpose and agenda. The workshop leaders trained the panelists on how to use the Bookmark method, the Ohio content standards and the test materials.

Bookmark procedure. Ordered-Item Booklets were provided to the panelists, who were asked to make a judgment about "the divide between items that a student at the threshold of a performance level (the minimally qualified student) should master from those items that are not necessary to master" (Mitzel et al., 2001, p. 254). Each panelist placed a bookmark on that page of the *Ordered-Item Booklet.*

Panelists placed their bookmarks with respect to a response probability (RP) judgment – the likelihood, for example, that a just barely Proficient student is likely to respond successfully to an item. RP selection has implications both for how panelists construe a just barely Proficient student and the consequences of placing a bookmark in a given location in the *Ordered-Item Booklet*. RP values adopted for each standard-setting panel were guided by several considerations, including a desire for consistency in procedures across grades and subjects, as well as test difficulty.

Panelists began the standard-setting process by placing bookmarks for the Proficient cut score in two rounds. In round 1, panelists were asked to place the bookmark on the page where they would expect, for example, one-half of students who are just barely Proficient to respond successfully and to record the page number of the bookmark on the rating form.

Table leaders opened Round 2 with a discussion on agreement data feedback and estimated impact data. Panelists were first provided feedback on individual panelist-recommended Proficient cut scores. Table-level discussions focused on the lowest and highest recommended Proficient cut scores and the table's median score. Panelists were also provided median table scores for each table within the panel. Following table-level discussions of individual panelist placements, the discussion moved to the panel level to discuss differences across tables. In addition, panelists were provided with a lookup table of impact data that indicated, for each page in the *Ordered-Item Booklet*, the percentage of students who would meet or exceed any given performance standard if a bookmark were placed on that page. After completing their review and discussions of the agreement and impact data, panelists again placed their bookmarks.

Following placement of the Proficient cut scores in round 2, panelists recommended cut scores for Basic, Accelerated and Advanced performance standards. Cut scores for these performance standards were also made in two rounds, as described above, but panelists recommended all three cut scores in each round. Following final placement of all recommended cut scores, ODE

submitted the recommendations of each panel was to the Ohio State Board of Education for their review and adoption.

Equating Procedures

This section briefly describes the process we use to estimate field test item parameters to develop the achievement test scales and how items on subsequent operational test forms are linked to item parameters in the original field test administration. Following each operational administration of the Ohio Achievement Tests, AIR produces an Early Return Technical Report that fully documents the procedures employed to identify the equating samples, calibration and equating of operational test items, construction of conversion tables, and projection of operational test results.

Ohio employs a pre-equating strategy to facilitate the development of nearly equivalent operational test forms across administrations. Because item parameter estimates may shift across test administrations, it can be useful to augment the pre-equating methodology with a method for evaluating the applicability of the field test-derived item parameter estimates to each operational assessment. This strategy allows AIR to modify item parameter estimates in the operational assessment in the event that field test values no longer demonstrate acceptable levels of fit to the measurement model. To accomplish this, for each operational assessment, AIR typically identifies an early return sample of schools that comprises at least 10,000 students. Schools assigned to the early return sample are simply designated for early document processing and scoring. In all other respects, test materials from the early return samples are processed in the same way as the remainder of the test population.

Historically, parameter estimates for all items included in operational forms are derived from analyses of the initial field test forms. To simultaneously maximize the number of field-tested items while minimizing the number of students participating in the field tests, a randomly equivalent groups equating design was employed for grades 3, 4 and 5 reading; grade 3 mathematics; and grade 4 writing. Test forms were spiraled within class to randomly assign test forms to students participating in the field test administration. Assuming that students responding to each test form represent randomly equivalent samples, parameters for items were estimated across forms by holding mean student performance across forms constant.

Beginning with the Fall 2004 field test administration, ODE adopted a common item equating design for item parameter estimation. To implement this, AIR developed a common item block design to augment the equivalent groups design. Parameter estimates for the grades 6, 7 and 8 reading and grades 4, 5, 6, 7 and 8 mathematics field tests reflected the common item block equating design.

Item parameters estimated from the field test administration are referred to as *item bank parameter estimates*. Procedures outlined in the steps below are used to place student responses from each operational test administration onto the Ohio achievement scale for each grade and subject test.

Step 1. *Centering on the bank*. Field-testing for grades 3, 4, and 5 reading, grade 3 mathematics, and grade 4 writing employed a randomly equivalent groups design to maximize the number of field-tested items with the fewest students. As a consequence of this design, within each grade and subject each of the field test forms were centered on theta. By initially centering on theta, item parameters across all forms are placed on the same scale, but do not have a mean value equal to zero. Therefore, prior to the first operational form administration, the bank of items was re-centered so that the pool of items had an average value of zero. This was accomplished by doing the following:

a) For each item in the bank, simple averages of the item difficulty values were computed.

b) For items appearing in multiple forms, weighted averages of the item difficulty values obtained from the field test (weighted by the inverse of the measurement error variances of the difficulty values) were computed and we used these weighted averages as the item difficulty values for those items.

c) The average of these averages was used as the *centering constant*. After applying this *centering constant* to the item difficulty values in the bank (by subtracting this constant from all of the item difficulty values), the bank had an average difficulty equal to zero.

Step 2: *Centering on the first operational form*. For all grades and subjects, it is ODE's policy to re-center the bank after the first operational administration of the test so that the initial administration of the test has a mean difficulty of zero. This requires a re-centering of the bank following the first operational administration.

a) Item difficulty values for the first operational administration were estimated based on the early return sample using WINSTEPS®. WINSTEPS employs a joint maximum likelihood approach to estimation (JMLE), which jointly estimates the person and item parameters.

b) Next, the Ohio linking procedure was applied to equate the first operational administration item difficulty values with the weighted averaged item difficulty values from the field test administration. This resulted in the identification of a *linking constant*. By subtracting this constant from all item difficulty values determined during the first operational test administration, the items on the first operational test and were put on the same scale as those on the field tests.

c) The simple average of all the item difficulty values in the first operational test form were then computed. The average of these averages yields the new first operational administration *centering constant*.

d) After applying the first operational administration *centering constant*, by subtracting the *centering constant* from all of the item difficulty values in the bank, all bank items are centered on the first operational form, which has an average difficulty equal to zero.

Subsequent operational test forms do not need to have re-centered item parameters, unlike the first operational test administrations. Nevertheless, for each operational administration, items must be linked to the item bank parameters. The linking procedure is described in the following section.

Estimation of IRT models. AIR uses Masters' (1982) Partial Credit Model, an extension of the one parameter Rasch model that allows for graded responses, to estimate item parameters for the Ohio Achievement Tests.

A stepwise deletion procedure was then used to calculate the linking constant needed to bring the set of operationally administered test items back to the reference scale. Following this procedure, the linking constant was first applied to bring the items back to the reference scale, and then examine the parameter estimates of anchor items to determine whether any exceed the .3 tolerance level for inclusion as anchor items. At each step, the item with the greatest difference between its linked and reference item parameter estimates was eliminated from the anchor set, provided the difference was greater than .3. We then computed a new linking constant, applied the linking constant to the test items, and then examined the resulting parameter estimates for the remaining anchor items to determine whether any exceeded the .3 tolerance level. This process was repeated until all remaining anchor items meet the tolerance level specifications.

Scaled scores and the Ohio rounding rule. Once the early return item parameters have been linked to the appropriate item bank reference scale, several steps are followed to transform raw scores to the Ohio Achievement Test reporting scale. The Ohio Achievement Test scaled scores represent a linear transformation of the Rasch ability estimates (theta scores), with the proficient cut score or performance standard set at a scaled score of 400. To transform student scores from the theta metric to the Ohio Achievement Test scale, the theta value associated with the bookmark page defining the proficiency performance standard (CutScore[theta]) was first identified. To determine the scaled scores associated with the other theta values, the following formula is implemented:

$$\text{Scale Score} = 400+(30*(\{\text{theta} - \text{CutScore[theta]}\} / \text{SD[theta]}) \qquad (1)$$

where 400 is the scaled score representing the proficiency standard cut score on the Ohio Achievement Tests, and 30 is the standard deviation of the Ohio Achievement Test scale. The "theta" represents any level of student ability on the operational form or *Ordered-Item Booklet*. The CutScore[theta] represents the theta that the panelists determined for the Proficient Level cut score from the *Ordered-Item Booklet*. The SD[theta] represents the standard deviation of all the thetas, or logit values. Cut scores for the Basic, Accelerated, and Advanced performance standards can then be located on the scale by finding the theta value associated with each page in the *Ordered-Item Booklet* corresponding to the relevant performance standard. Table 1 presents the theta to scaled score linear transformation equations for grade 3 reading.

Table 1.
Theta to Scaled Score Linear Transformation Equations

Ohio Achievement Test	Linear Transformation Equation

Grade 3 Reading

$$SS = 400 + 30 \times \frac{\theta - 0.876834189407528}{1.19399161773822}$$

For score reporting, if the exact theta value corresponding to a performance standard does not appear in the operational form, then the ODE implements a rounding rule to determine the placement of the cut scores on the operational form. To implement the Ohio rounding rule, we first identify the two theta values closest to the performance standard theta (one above and one below) and select the one nearest to the standard set by panelists. If the theta nearest to the performance standard is below the standard (or smaller in value), then that theta is rounded up to the theta associated with the performance standard. If the nearest operational test theta is greater than the theta associated with the standard set by the panelists, then that theta is selected as the operational test cut score.

Operational Test Results

The remainder of this report summarizes the operational test results for the October 2008 administration of the Ohio Achievement Tests in grade 3 reading.

Parameter estimation, equating and scaling for the October 2008 administration employed an early return sample of Ohio public school students administered the grade 3 reading test. If a school was selected as an early return school, the school was asked to administer and return their testing materials on an expedited schedule. In all other respects, however, test materials from the early return sample are processed in the same way as the remainder of the test population. In this report we briefly summarize results from the the analysis of the early return sample data. A complete description of the early return sample and analysis results is provided in the Early Return Technical Report available from ODE. Table 2 provides the early return sample and final operational student counts. Assessment data for both the early return and final operational analyses included only Ohio public school students.

Table 2.
Counts for Early Return Sample and Final Operational Assessment – Ohio Public School Students Only

Operational Assessment	Early Return Sample Counts	Final Operational Assessment Student Counts
Grade 3 Reading	14140	130000

This section of the report is organized into the following parts:

Classical Item Analysis

Rasch/Item Response Theory analysis

Raw Score and Scaled Score Means, Standard Deviations and Frequency Distributions

Generalizability Study

Subscale Analysis.

Classical Item Analysis

Traditional item statistics for multiple-choice (MC) and constructed-response (CR) items were calculated in the operational test forms based on the early return samples. Appendix A presents item statistics for the operational items. Item statistics include the proportion of students falling into each score-point category (e.g., 0, 1 for multiple-choice items, 0, 1, 2 for short-answer items, and 0, 1, 2, 3, 4 for extended response items), as well as the proportion of students with omitted responses. Also presented are *p*-values and average item scores for MC items and CR items, respectively, along with adjusted item-test biserial/polyserial correlations, maximum point values and subscale information.

The tables in Appendix B provide the raw score and scaled score means, standard deviations, and standard errors of measurement, as well as the internal consistency reliability estimates for each test, based on all Ohio public school students. Appendix B also presents internal consistency estimates for gender and ethnic subgroups.

Table 3 compares item characteristics between the early return and field test samples for items in the operational form. The average difference between field test and operational test *p*-values was -.05, indicating that *p*-values generally decreased from field test to operational test administration, with students performing slightly less well on these items in the context of the fall operational administration. Because the Grade 3 Reading items were field-tested in the spring semester of the school year, this difference may simply reflect differences in student achievement between the fall and spring semester. Adjusted item-total correlations were also slightly lower for the operational test administration, with an average difference of -.02.

Table 3.
Comparing Operational Test Early Return Sample and Field Test Sample Item Statistics – Grade 3 Reading

Items	Possible Points	*p*-Values or Average Proportion of Total Points			Adjusted Biserial or Polyserial Correlation		
		Field Test	Operational Test	Difference	Field Test	Operational Test	Difference
1	1	0.97	0.91	-0.06	0.64	0.58	-0.06
2	1	0.95	0.91	-0.04	0.65	0.65	0.00
3	1	0.95	0.87	-0.08	0.72	0.65	-0.07
4	1	0.91	0.83	-0.08	0.67	0.63	-0.04
5	1	0.91	0.88	-0.03	0.49	0.50	0.01
6	1	0.71	0.65	-0.06	0.57	0.57	-0.00
7	1	0.93	0.93	0.00	0.73	0.69	-0.04
8	4	0.60	0.52	-0.09	0.54	0.56	0.02
9	1	0.81	0.80	-0.00	0.55	0.56	0.02
10	1	0.86	0.85	-0.01	0.75	0.74	-0.01
11	1	0.66	0.59	-0.07	0.49	0.44	-0.05
12	1	0.87	0.89	0.02	0.82	0.68	-0.14
13	2	0.56	0.42	-0.14	0.61	0.65	0.05
14	1	0.87	0.82	-0.05	0.74	0.72	-0.02
15	1	0.76	0.70	-0.06	0.51	0.54	0.03
16	1	0.88	0.83	-0.05	0.78	0.69	-0.09
17	4	0.56	0.50	-0.06	0.61	0.62	0.01
18	1	0.86	0.80	-0.06	0.79	0.73	-0.06
19	1	0.88	0.80	-0.08	0.66	0.65	-0.01
20	1	0.92	0.89	-0.03	0.61	0.63	0.02
21	1	0.71	0.67	-0.04	0.59	0.49	-0.09
22	2	0.41	0.36	-0.05	0.61	0.58	-0.03
23	1	0.72	0.65	-0.07	0.73	0.72	-0.01
24	1	0.90	0.87	-0.03	0.66	0.70	0.04
25	1	0.28	0.26	-0.02	0.31	0.30	-0.01

26	1	0.69	0.64	-0.05	0.73	0.64	-0.09
27	1	0.89	0.83	-0.06	0.73	0.71	-0.02
28	4	0.52	0.48	-0.04	0.56	0.56	0.01
29	1	0.78	0.79	0.02	0.57	0.58	0.02
30	1	0.84	0.80	-0.04	0.65	0.64	-0.01
31	2	0.69	0.55	-0.14	0.72	0.72	0.00
32	1	0.79	0.75	-0.04	0.66	0.69	0.03
33	1	0.83	0.76	-0.07	0.73	0.71	-0.02
34	2	0.35	0.26	-0.09	0.59	0.59	0.00
35	1	0.82	0.78	-0.04	0.69	0.65	-0.04
36	1	0.75	0.68	-0.07	0.57	0.52	-0.05

* Change = Operational value – Field test value.

Item Response Theory Analysis

For pragmatic reasons, AIR uses Masters' (1982) partial credit model, an extension of the one parameter Rasch model that allows for graded responses, to estimate item parameters for the Ohio Achievement Tests. The principal advantage of the Rasch model is the resulting one-to-one correspondence between the number of correct responses and the scaled score produced by the model.

IRT model item parameters were estimated using WINSTEPS, publicly available IRT software from Mesa Press. WINSTEPS employs a joint maximum likelihood approach to estimation (JMLE), which jointly estimates the person and item parameters.

We applied Masters' Partial Credit Model to estimate the Rasch model parameters for the grade 3 reading test based on the early return sample. Appendix C presents the item statistics resulting from the free (unanchored) estimation of parameters for operational test items. The column "Num" refers to the item's order in the operational test booklet. The column "Score" is the sum of the points received by those responding to the question. "Count" is the number of students responding to the item. The item difficulty estimates are presented in the column "Measure," followed by the standard error of estimate and the infit and outfit fit statistics.

A stepwise deletion procedure is used to calculate the linking constant needed to bring the set of items back to the reference scale. Following this procedure, the first step is to use all the anchor items to calculate the linking constant necessary to bring the operational test items back to the reference scale. Linked early return and bank item parameter estimates are then compared for each anchor item to determine whether any exceeds the .3 tolerance level for inclusion as an anchor item. At each step, we eliminate from the anchor set the item with the greatest difference between its linked and reference item parameter estimates, provided the difference is greater than .3. We then compute and apply a new linking constant to the test items, and examine the parameter estimates for remaining anchor items to determine whether any exceed the .3 tolerance level. We repeat this process until all remaining anchor items meet the tolerance level specifications. Appendix D provides comparisons between item parameter estimates that result from the iterative application of the linking constant to the early return sample with parameter

estimates obtained from the field test sample and the unanchored calibration and equating sample.

Table 4 summarizes the number of items used to link the October 2008 operational form to the grade 3 reading Ohio Achievement Test scale. Following application of the .3 rule, 75% of operational items were used to identify the linking constant.

Table 4.
Summary of Items Used to Link October 2008 Operational Test to Ohio Achievement Test Scales

	Summary of Linking Items									
Test	**Content Standard**					**MC Items**	**CR Items**	**Linking Items**	**Total Items**	**% Linking Items**
Reading	**AV**	**IT**	**LT**	**RP**						
G3R	6	8	6	7		21	6	27	36	75%

Appendix E presents the final transformations of raw scores to Rasch ability estimate (in logit measure) to scaled score based on the final set of anchor items. This table also lists the error of estimation for each value, as well as the proficiency level associated with each score point. Extreme scores (i.e., 0 and perfect) are estimated using a linear extrapolation procedure in which theta(0) = 2*theta(1)-theta(2) and theta(n) = 2*theta(n-1)-theta(n-2). This procedure produces values very close to those obtained by adding or subtracting .5 from zero or perfect scores, respectively, as recommended by Berkson (as cited in Linacre, 2004).

Appendix F represents the ability parameters associated with the cut scores for previous and current operational test administrations. Table 5 provides the average item difficulty and student ability estimates across all available fall operational administrations. As the table indicates, student performance increased in each operational administration.

Table 5.
Average Item Difficulty and Student Ability Estimates across Test Administrations Based on Early Return Samples

Test Administration	Form Difficulty	Avg. Student Ability - Theta	Avg. Student Ability - Scaled
October 2003	**0.00**	**0.91**	**400.8**
October 2004	**-0.12**	**0.93**	**401.2**
October 2005	**-0.24**	**1.02**	**403.7**
October 2006	**0.05**	**1.10**	**405.5**
October 2007	**0.18**	**1.17**	**407.3**
October 2008	**-0.17**	**1.12**	**406.0**

Appendix G presents classification consistency estimates at each of the proficiency classification cut scores on the operational tests and subscale scores. Classification consistency indexes the agreement between the classification resulting from students' observed scores and the classification resulting from scores as projected from a hypothetical independent administration

of a parallel test form. Classification consistency estimates are derived following Huynh's (1979) use of the beta-binomial model. Kappa provides an index of classification consistency that is corrected for chance levels of agreement.

Appendix H presents the percentage of students scoring at each proficiency level, both overall and disaggregated by gender and ethnicity. The tables in Appendix I present the percentage of students scoring at or above each performance standard.

Raw and Scaled Scores: Descriptive Statistics

Raw score and scaled score means and standard deviations for public and non-public schools are presented in Appendix J and Appendix K, respectively. For public schools, raw score and scaled score cumulative frequency distributions for the total population and by gender are presented in Appendix L, with distributions by ethnicity presented in Appendix M. The cumulative frequency distribution for non-public school students is presented in Appendix N.

Interrater Analysis: G-Study

The goal of the Generalizability Study (G-study) is to identify undesirable judging or rating behavior in the scoring of constructed-response items. Ten percent (10%) of student papers were double-scored by approximately 100 trained raters. The reader is referred to the October 2008 Generalizability Study Technical Report for a complete description of the G-study methods and results.

The primary technical difficulty in conducting a multi-facet G-study is that the 10 percent of double-scored CR items are randomly selected from across the pool of item responses, not students. In other words, rather than randomly selecting 10% of students and double scoring all CR items for those students, a random sample of 10% of item responses were selected for double scoring, so that for some students only one item response may have been double-scored, while for other students, two items may have been double-scored, and so on. With existing models, the best analysis one can perform is a traditional reliability analysis of ratings for each individual item.

To obtain maximum efficiency, we employed the Maximal Fully Crossed method (Antal, Johnson, & Antal, 2004). By using subsets from the original pool of double-scored items, the Maximal Fully Crossed method allows for the execution of a more complex G-study that includes more than one facet, permitting the estimation of more variance components and multiple G-coefficients. Two facet designs, based on item pairs, triplets, and quartets, were used to estimate variance components. The model's main elements are the true score, a random variable defined over the universe of persons, items and ratings. It must be emphasized that *rater* reliability cannot be established due to the fact that papers are randomly assigned to raters, and that rater pairs were not selected systematically for blocks of items. From a design point of view, the question addressed is whether the first and second ratings are consistent, while the performance of the individual scorer falls out of the scope of the present investigation.

Generalizability analyses were carried out on single items, item pairs, item triplets, and item quartets for constructed-response items. Appendix O presents interrater analysis results for the

single facet and each of the two-facet designs. For each model, the last column in Table O1 presents the average generalizability coefficients for constructed response items.

To better understand the behavior of the variance components relative to one another, Table O1 also presents the variance component means for the one-facet and two-facet designs. For the one-facet analysis, the variance due to ratings is essentially zero. For the two-facet analysis, all variance components that incorporate rating performance (excluding the error component) are also nearly zero. The person-item interaction component causes the largest variability, especially in generalizability studies of constructed-response items. Across subset types, the person-item interaction component was consistent, ranging from $\sigma^2_{(pi)} = 0.079$ to $\sigma^2_{(pi)} = 0.081$. The person variance component was the second largest component in the OAT constructed-response items. The person variance component was also consistent across subset types, ranging from $\sigma^2_{(p)} = 0.042$ to $\sigma^2_{(p)} = 0.046$ in the two-facet designs. Item variances were equivalent across subset types ($\sigma^2_{(i)} = 0.012$). The error component was also nearly equivalent across subset types, ranging from $\sigma^2_{(pir,e)} = 0.006$ to $\sigma^2_{(pir,e)} = 0.007$.

Subscale Analysis

Appendix P presents the coefficient alpha reliability, the passing band and the maximum possible score for each subscale. Raw score frequency distributions for each of the subscales, for public school students only, are presented in Appendix Q.

The tables in Appendix R present the intercorrelations among the content standard subscales for the Ohio achievement tests. We note that the intercorrelations among subscales is quite high, and these values are likely attenuated due to the relatively moderate to low internal consistency reliability estimates for these subscales. It is therefore important to use caution when interpreting differences between content standard scores.

References

Antal, J., Johnson, E.G., & Antal, T. (2004). *Generalizability analysis with maximal fully crossed subsets*. Unpublished manuscript.

Huynh, H. (1979). Statistical inference for two reliability indices in mastery testing based on the beta-binomial model. *Journal of Educational Statistics*, *4*, 231-246.

Linacre, J.M. (2004). *A user's guide to WINSTEPS*. Chicago: MESA Press.

Longford, N.T. (1994). *Summarizing rater reliability and departure from unidimensionality in tests with multiple subjectively rated responses*. Unpublished manuscript.

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

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Erlbaum.

APPENDIX A

Early Return Sample Item Statistics

Table A1. Early Return Sample Item Statistics – Grade 3 Reading

Position	Max score.	Type	Subscale	r_b	*p*-Val/ Avg Scr	Proportion of Students at Each Score Point						
						Omit	Not Reached	0	1	2	3	4
1	1	MC	AV	0.58	0.91	0.00	0.00	0.09	0.91			
2	1	MC	AV	0.65	0.91	0.00	0.00	0.08	0.91			
3	1	MC	AV	0.65	0.87	0.01	0.00	0.12	0.87			
4	1	MC	AV	0.63	0.83	0.00	0.00	0.17	0.83			
5	1	MC	AV	0.50	0.88	0.00	0.00	0.12	0.88			
6	1	MC	IT	0.57	0.65	0.00	0.00	0.35	0.65			
7	1	MC	RP	0.69	0.93	0.00	0.00	0.06	0.93			
8	4	EA	IT	0.56	0.52	0.01	0.00	0.09	0.19	0.32	0.30	0.08
9	1	MC	IT	0.56	0.80	0.00	0.00	0.20	0.80			
10	1	MC	IT	0.74	0.85	0.00	0.00	0.14	0.85			
11	1	MC	IT	0.44	0.59	0.00	0.00	0.41	0.59			
12	1	MC	RP	0.68	0.89	0.01	0.00	0.09	0.89			
13	2	SA	RP	0.65	0.42	0.01	0.00	0.45	0.25	0.29		
14	1	MC	AV	0.72	0.82	0.00	0.00	0.17	0.82			
15	1	MC	LT	0.54	0.70	0.00	0.00	0.29	0.70			
16	1	MC	RP	0.69	0.83	0.01	0.00	0.17	0.83			
17	4	EA	RP	0.62	0.50	0.03	0.00	0.19	0.14	0.22	0.26	0.16
18	1	MC	LT	0.73	0.80	0.01	0.00	0.19	0.80			
19	1	MC	LT	0.65	0.80	0.01	0.00	0.18	0.80			
20	1	MC	AV	0.63	0.89	0.00	0.00	0.10	0.89			
21	1	MC	LT	0.49	0.67	0.02	0.00	0.31	0.67			
22	2	SA	RP	0.58	0.36	0.02	0.00	0.42	0.40	0.16		
23	1	MC	RP	0.72	0.65	0.00	0.00	0.35	0.65			
24	1	MC	AV	0.70	0.87	0.01	0.00	0.12	0.87			
25	1	MC	AV	0.30	0.26	0.00	0.00	0.74	0.26			
26	1	MC	LT	0.64	0.64	0.01	0.00	0.34	0.64			
27	1	MC	RP	0.71	0.83	0.02	0.00	0.15	0.83			
28	4	EA	LT	0.56	0.48	0.02	0.01	0.15	0.19	0.29	0.22	0.12
29	1	MC	LT	0.58	0.79	0.03	0.01	0.18	0.79			
30	1	MC	IT	0.64	0.80	0.01	0.01	0.20	0.80			
31	2	SA	RP	0.72	0.55	0.03	0.01	0.35	0.16	0.47		
32	1	MC	AV	0.69	0.75	0.01	0.01	0.24	0.75			
33	1	MC	IT	0.71	0.76	0.01	0.01	0.24	0.76			
34	2	SA	IT	0.59	0.26	0.03	0.01	0.59	0.26	0.13		
35	1	MC	IT	0.65	0.78	0.02	0.01	0.20	0.78			
36	1	MC	RP	0.52	0.68	0.00	0.02	0.32	0.68			

Note: Type: MC = Multiple choice, EA = Extended Answer, SA = Short Answer; Subscale: AV = Vocabulary; RP=Reading Process; IT=Information Text; LT=Literary Text

APPENDIX B

Operational Test Summary Statistics

Table B1. Operational Test Summary Statistics

Test Grade / Subject	N-count	Max Obtained Raw Score	Raw Score Mean	Raw Score Standard Deviation	Raw Score SEM	Max Scaled Score	Scaled Score Mean	Scaled Score Standard Deviation	Scaled Score SEM	Reliability
Grade 3 Reading	130000	49	31.76	9.64	3.06	518	406.95	31.66	10.06	0.90

Table B2.1. Operational Internal Consistency Estimates for Subgroups – Grade 3 Reading

Gender/Ethnicity	Sample Size	Internal Consistency Reliability Estimates (Cronbach's α) Total Reading	AV	IT	LT	RP
All Public School Students	**130000**	**0.90**	**0.70**	**0.70**	**0.62**	**0.75**
Gender						
Female	**63448**	**0.89**	**0.68**	**0.69**	**0.61**	**0.74**
Male	**66032**	**0.90**	**0.71**	**0.71**	**0.63**	**0.76**
Ethnicity						
American Indian	**175**	**0.92**	**0.75**	**0.73**	**0.67**	**0.79**
Asian/Pacific Islander	**2261**	**0.89**	**0.68**	**0.69**	**0.61**	**0.73**
Black/African American	**19881**	**0.90**	**0.70**	**0.68**	**0.63**	**0.75**
Hispanic	**3741**	**0.90**	**0.70**	**0.71**	**0.62**	**0.75**
White	**94983**	**0.89**	**0.66**	**0.68**	**0.60**	**0.73**
Multi-Ethnic	**5604**	**0.90**	**0.69**	**0.69**	**0.62**	**0.74**
Other	**1261**	**0.90**	**0.72**	**0.71**	**0.63**	**0.76**

Note. AV – Acquisition of Vocabulary; IT – Informational Text; LT – Literary Text; RP – Reading Process.

APPENDIX C

Early Return Sample Item Calibration

Table C1. Early Return Sample WINSTEPS Item Calibrations – Grade 3 Reading

Num	Score	Count	Measure	SE	Infit	Infit zstd	Outfit	Outfit zstd
1	12810.60	14131.20	-2.78	0.03	0.97	-1.51	0.95	-1.03
2	12912.80	14128.80	-2.89	0.03	0.94	-2.62	0.72	-6.13
3	12267.20	14131.30	-2.34	0.03	0.95	-3.07	0.78	-6.12
4	11685.90	14132.20	-1.96	0.02	0.96	-2.79	0.86	-4.64
5	12428.40	14131.50	-2.46	0.03	1.04	2.15	1.11	2.73
6	9184.80	14130.20	-0.78	0.02	1.01	1.03	0.98	-1.06
7	13207.20	14131.10	-3.22	0.04	0.90	-3.77	0.62	-7.42
8	29211.40	14131.20	0.00	0.01	1.26	9.90	1.26	9.90
9	11280.40	14127.60	-1.73	0.02	1.02	1.20	1.04	1.38
10	12026.00	14127.20	-2.17	0.03	0.88	-8.16	0.65	-9.90
11	8277.10	14126.30	-0.43	0.02	1.13	9.90	1.17	9.90
12	12611.30	14127.30	-2.61	0.03	0.90	-5.10	0.82	-4.35
13	11717.80	14127.30	0.42	0.01	1.01	0.90	0.97	-1.69
14	11630.90	14124.50	-1.93	0.02	0.88	-8.60	0.74	-9.12
15	9951.00	14119.40	-1.10	0.02	1.05	4.47	1.04	2.28
16	11656.50	14120.20	-1.94	0.02	0.92	-5.96	0.76	-8.27
17	28353.70	14117.80	0.08	0.01	1.31	9.90	1.36	9.90
18	11284.20	14113.30	-1.73	0.02	0.87	-9.90	0.72	-9.90
19	11328.80	14111.40	-1.76	0.02	0.95	-4.13	0.86	-5.23
20	12608.90	14106.90	-2.62	0.03	0.93	-3.74	1.09	1.98
21	9492.80	14083.80	-0.91	0.02	1.10	9.46	1.07	4.16
22	10132.20	14082.20	0.80	0.01	1.02	1.64	1.00	0.10
23	9144.90	14076.60	-0.77	0.02	0.86	-9.90	0.78	-9.90
24	12280.90	14075.80	-2.38	0.03	0.91	-5.56	0.68	-9.13
25	3612.40	14070.60	1.35	0.02	1.11	9.90	1.81	9.90
26	9019.30	14068.40	-0.72	0.02	0.93	-7.21	0.90	-6.67
27	11655.70	14066.40	-1.96	0.02	0.90	-7.28	0.74	-9.09
28	27086.50	14054.20	0.14	0.01	1.35	9.90	1.36	9.90
29	11049.00	14030.40	-1.64	0.02	1.02	1.20	0.96	-1.42
30	11159.10	14006.00	-1.71	0.02	0.97	-2.46	0.80	-8.01
31	15290.60	13988.40	-0.15	0.01	0.99	-0.45	0.98	-1.18
32	10521.30	13979.30	-1.39	0.02	0.91	-7.59	0.78	-9.90
33	10575.90	13962.40	-1.42	0.02	0.90	-8.69	0.74	-9.90
34	7109.20	13956.70	1.29	0.01	0.95	-4.36	0.89	-5.40
35	10902.90	13935.20	-1.59	0.02	0.94	-4.63	0.83	-6.78
36	9497.80	13872.90	-0.96	0.02	1.06	6.10	1.02	0.88
Mean	**12360.15**	**14077.94**	**-1.28**	**0.02**	**0.99**	**-1.11**	**0.94**	**-2.48**
SD	**5247.73**	**67.95**	**1.18**	**0.01**	**0.12**	**6.30**	**0.24**	**6.55**

APPENDIX D

Comparing Field Test, Early Return and Linked Early Return Sample Item Parameter Estimates

Table D1. Comparing Field Test, Operational, and Linked Operational Item Parameter Estimates – Grade 3 Reading

Item Position	Operational	Average Bank	Step 1 Adjusted	Step 2 Adjusted	Step 3 Adjusted	Step 4 Adjusted	Step 5 Adjusted	Step 6 Adjusted	Step 7 Adjusted	Step 8 Adjusted	Step 9 Adjusted	Step 10 Adjusted	Final Linked
1	-2.78	-2.46	-1.73	-1.71									-1.67
2	-2.89	-1.89	-1.84	-1.82	-1.79	-1.81	-1.80	-1.81	-1.80	-1.78	-1.77	-1.78	-1.78
3	-2.34	-2.02	-1.29										-1.23
4	-1.96	-1.34	-0.91	-0.89	-0.86	-0.88	-0.87	-0.88					-0.85
5	-2.46	-1.45	-1.41	-1.39	-1.36	-1.38	-1.37	-1.38	-1.37	-1.35	-1.34	-1.35	-1.35
6	-0.78	0.26	0.27	0.29	0.32	0.30	0.31	0.30	0.31	0.33	0.34	0.33	0.33
7	-3.22	-1.60	-2.17	-2.15	-2.12	-2.14	-2.13						-2.11
8	0.00	1.04	1.06	1.08	1.10	1.08	1.10	1.08	1.10	1.11	1.12	1.11	1.11
9	-1.73	-0.66	-0.68	-0.66	-0.63	-0.65	-0.64	-0.65	-0.64	-0.62	-0.61	-0.62	-0.62
10	-2.17	-0.86	-1.12	-1.10	-1.07	-1.09	-1.08	-1.09	-1.08	-1.06	-1.05	-1.06	-1.06
11	-0.43	0.31	0.62	0.64	0.67	0.65	0.66	0.65	0.66	0.68			0.68
12	-2.61	-0.86	-1.56	-1.54	-1.51								-1.50
13	0.42	0.90	1.47	1.49	1.51	1.49							1.52
14	-1.93	-1.22	-0.88	-0.86	-0.83	-0.85	-0.84	-0.85	-0.84				-0.82
15	-1.10	0.07	-0.05	-0.03	0.00	-0.02	-0.01	-0.02	-0.01	0.01	0.02	0.01	0.01
16	-1.94	-0.84	-0.89	-0.87	-0.84	-0.86	-0.85	-0.86	-0.85	-0.83	-0.82	-0.83	-0.83
17	0.08	1.24	1.13	1.15	1.17	1.15	1.17	1.15	1.17	1.18	1.19	1.18	1.18
18	-1.73	-0.63	-0.68	-0.66	-0.63	-0.65	-0.64	-0.65	-0.64	-0.62	-0.61	-0.62	-0.62
19	-1.76	-0.88	-0.71	-0.69	-0.66	-0.68	-0.67	-0.68	-0.67	-0.65	-0.64	-0.65	-0.65
20	-2.62	-1.35	-1.57	-1.55	-1.52	-1.54	-1.53	-1.54	-1.53	-1.51	-1.50	-1.51	-1.51
21	-0.91	0.23	0.14	0.16	0.19	0.17	0.18	0.17	0.18	0.20	0.21	0.20	0.20
22	0.80	1.69	1.85	1.87	1.89	1.87	1.89	1.87	1.89	1.90	1.91	1.90	1.90
23	-0.77	0.30	0.28	0.30	0.33	0.31	0.32	0.31	0.32	0.34	0.35	0.34	0.34
24	-2.38	-1.00	-1.33	-1.31	-1.28	-1.30	-1.29	-1.30	-1.29	-1.27	-1.26	-1.27	-1.27
25	1.35	2.53	2.40	2.42	2.45	2.43	2.44	2.43	2.44	2.46	2.47	2.46	2.46
26	-0.72	0.38	0.33	0.35	0.38	0.36	0.37	0.36	0.37	0.39	0.40	0.39	0.39
27	-1.96	-0.78	-0.91	-0.89	-0.86	-0.88	-0.87	-0.88	-0.87	-0.85	-0.84	-0.85	-0.85
28	0.14	1.29	1.20	1.22	1.24	1.22	1.24	1.22	1.24	1.25	1.26	1.25	1.25
29	-1.64	-0.22	-0.59	-0.57	-0.54	-0.56	-0.55	-0.56	-0.55	-0.53	-0.52		-0.53
30	-1.71	-0.47	-0.66	-0.64	-0.61	-0.63	-0.62	-0.63	-0.62	-0.60	-0.59	-0.60	-0.60
31	-0.16	0.73	0.90	0.92	0.94	0.92	0.94	0.92	0.94	0.95	0.96	0.95	0.95
32	-1.39	-0.11	-0.34	-0.32	-0.29	-0.31	-0.30	-0.31	-0.30	-0.28	-0.27	-0.28	-0.28
33	-1.42	-0.44	-0.37	-0.35	-0.32	-0.34	-0.33	-0.34	-0.33	-0.31	-0.30	-0.31	-0.31
34	1.29	2.23	2.34	2.36	2.39	2.37	2.38	2.37	2.38	2.40	2.41	2.40	2.40
35	-1.59	-0.36	-0.54	-0.52	-0.49	-0.51	-0.50	-0.51	-0.50	-0.48	-0.47	-0.48	-0.48
36	-0.96	0.15	0.09	0.11	0.14	0.12	0.13	0.12	0.13	0.15	0.16	0.15	0.15
Mean	**-1.28**	**-0.22**	**-0.22**	**-0.17**	**-0.11**	**-0.08**	**-0.11**	**-0.07**	**-0.02**	**0.02**	**0.01**	**0.02**	**-0.17**
Std. Dev.	**1.18**	**1.17**	**1.18**	**1.18**	**1.17**	**1.16**	**1.15**	**1.10**	**1.11**	**1.12**	**1.13**	**1.15**	**1.18**
Constants			**-1.05**	**-1.07**	**-1.10**	**-1.08**	**-1.09**	**-1.08**	**-1.09**	**-1.11**	**-1.12**	**-1.11**	

APPENDIX E

Raw Score to Scaled Score Conversion Table

Table E1. Conversion Table – Grade 3 Reading

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
0	-5.01	1.45	252	1	Limited
1	-4.27	1.02	271	1	Limited
2	-3.53	0.74	289	1	Limited
3	-3.08	0.62	301	1	Limited
4	-2.74	0.55	309	1	Limited
5	-2.47	0.50	316	1	Limited
6	-2.24	0.47	322	1	Limited
7	-2.03	0.44	327	1	Limited
8	-1.85	0.42	332	1	Limited
9	-1.68	0.40	336	1	Limited
10	-1.52	0.39	340	1	Limited
11	-1.37	0.38	344	1	Limited
12	-1.23	0.37	347	1	Limited
13	-1.10	0.36	350	1	Limited
14	-0.97	0.36	354	1	Limited
15	-0.84	0.35	357	1	Limited
16	-0.72	0.35	360	1	Limited
17	-0.60	0.34	363	1	Limited
18	-0.49	0.34	366	1	Limited
19	-0.37	0.34	369	1	Limited
20	-0.26	0.33	371	1	Limited
21	-0.15	0.33	374	1	Limited
22	-0.04	0.33	377	1	Limited
23	0.06	0.33	380	1	Limited
24	0.17	0.33	382	1	Limited
25	0.28	0.33	385	2	Basic
26	0.38	0.33	388	2	Basic
27	0.49	0.33	390	2	Basic
28	0.60	0.33	393	2	Basic
29	0.70	0.33	396	2	Basic
30	0.81	0.33	398	2	Basic
31	0.93	0.34	401	3	Proficient
32	1.04	0.34	404	3	Proficient
33	1.16	0.34	407	3	Proficient
34	1.28	0.35	410	3	Proficient
35	1.47	0.36	415	4	Accelerated
36	1.53	0.36	416	4	Accelerated
37	1.67	0.37	420	4	Accelerated
38	1.81	0.38	423	4	Accelerated
39	1.96	0.40	427	4	Accelerated
40	2.15	0.41	432	5	Advanced
41	2.30	0.43	436	5	Advanced
42	2.50	0.45	441	5	Advanced
43	2.71	0.48	446	5	Advanced
44	2.95	0.51	452	5	Advanced
45	3.24	0.56	459	5	Advanced

Raw Score	Theta	S.E.	Scaled Score	Performance Level	Proficiency
46	3.59	0.63	468	5	Advanced
47	4.06	0.75	480	5	Advanced
48	4.82	1.04	499	5	Advanced
49	5.59	1.46	518	5	Advanced

APPENDIX F

Ability Measures at Cut Scores

Table F1. Ability Measures at Raw Score Cuts

	Basic			Proficient			Accelerated			Advanced		
	Raw Score	**Theta***	**Scaled Score***	**Raw Score**	**Theta**	**Scaled Score**	**Raw Score**	**Theta**	**Scaled Score**	**Raw Score**	**Theta**	**Scaled Score**
G3R Standard Setting Values		**0.27**	**385**		**0.88**	**400**		**1.47**	**415**		**2.15**	**432**
G3R October 2003 Early Return Sample	25	0.24	384	33	0.92	401	38	1.43	414	43	2.17	432
G3R March 2004 Early Return Sample	24	0.24	384	31	0.85	399	37	1.45	414	42	2.14	432
G3R October 2004 Early Return Sample	26	0.31	386	32	0.91	401	37	1.50	416	41	2.09	431
G3R March 2005 Early Return Sample	26	0.24	384	32	0.87	400	38	1.52	416	43	2.17	432
G3R October 2005 Early Return Sample	25	0.25	384	31	0.86	400	37	1.52	416	41	2.07	430
G3R March 2006 Early Return Sample	25	0.29	385	31	0.92	401	36	1.49	416	41	2.19	433
G3R October 2006 Early Return Sample	24	0.24	384	31	0.87	400	37	1.47	415	42	2.14	432
G3R May 2007 Early Return Sample	25	0.29	385	31	0.89	400	36	1.43	414	41	2.10	431
G3R October 2007 Early Return Sample	23	0.31	386	29	0.92	401	34	1.47	415	39	2.15	432
G3R May 2008 Early Return Sample	26	0.30	386	33	0.92	401	38	1.41	413	44	2.24	434
G3R October 2008 Early Return Sample	25	0.28	385	31	0.93	401	35	1.40	413	40	2.13	431

Note. Theta and Scaled Score values reported in this table do not reflect the Ohio Rounding Rule.

APPENDIX G

Early Return Sample Proficiency Classification Consistency Rates

Table G1. Early Return Sample Proficiency Classification Consistency Rates – Grade 3 Reading

Proficiency Level Classification	Consistency	$\kappa_{Consistency}$
Grade 3 Reading – Total		
Basic	.89	.71
Proficient	.86	.72
Accelerated	.86	.72
Advanced	.89	.68
G3R – Acquisition of Vocabulary		
At	.81	.40
Above	.70	.41
G3R – Reading Process		
At	.85	.58
Above	.80	.60
G3R – Informational Text		
At	.81	.45
Above	.74	.47
G3R – Literary Text		
At	.81	.47
Above	.76	.47

APPENDIX H

Operational Public School Percent of Students At Each Performance Level

Table H1. Percentage of Students At Each Test Performance Level – Grade 3 Reading

G3R	Limited	Basic	Proficient	Accelerated	Advanced
Overall	23.70	15.72	13.71	21.82	25.05
Female	20.81	15.24	14.24	22.57	27.14
Male	26.35	16.14	13.22	21.14	23.15
American Indian	30.86	16.00	16.00	17.71	19.43
Asian/Pacific Islander	12.25	11.15	10.92	21.58	44.10
Black/African American	44.60	18.80	13.04	14.66	8.89
Hispanic	40.34	18.74	12.32	15.77	12.83
White	18.55	14.91	13.93	23.63	28.97
Multi-Ethnic	26.16	16.90	14.22	22.14	20.57
Other	30.06	17.53	13.72	18.72	19.98

APPENDIX I

Operational Public School Percent of Students At or Above Each Performance Level

Table I1. Percentage of Students At or Above Each Performance Level – Grade 3 Reading

G3R	Basic	Proficient	Accelerated	Advanced
Overall	76.30	60.59	46.87	25.05
Female	79.19	63.94	49.71	27.14
Male	73.65	57.51	44.29	23.15
American Indian	69.14	53.14	37.14	19.43
Asian/Pacific Islander	87.75	76.60	65.68	44.10
Black/African American	55.40	36.60	23.56	8.89
Hispanic	59.66	40.92	28.60	12.83
White	81.45	66.54	52.60	28.97
Multi-Ethnic	73.84	56.94	42.72	20.57
Other	69.94	52.42	38.70	19.98

APPENDIX J

Operational Public School Summary Statistics by Gender and Ethnicity

Table J1. Operational Public School Summary Statistics by Gender and Ethnicity: Grade 3 Reading

	N	α	Raw Score		Scaled Score	
			Mean	SD	Mean	SD
Total	130000	0.90	31.76	9.64	406.95	31.66
Female	63448	0.89	32.53	9.32	409.45	30.99
Male	66032	0.90	31.06	9.86	404.68	32.08
American Indian	175	0.92	29.24	10.89	398.84	35.60
Asian/Pacific Islander	2261	0.89	36.01	8.76	422.40	32.17
Black/African American	19881	0.90	25.99	9.88	388.34	30.30
Hispanic	3741	0.90	27.30	10.03	392.62	31.24
White	94983	0.89	33.17	9.02	411.48	30.23
Multi-Ethnic	5604	0.90	30.93	9.51	404.02	30.63
Other	1261	0.90	29.90	10.03	401.00	32.40

APPENDIX K

Operational Non-Public School Summary Statistics by Gender and Ethnicity

Table K1. Operational Non-Public School Summary Statistics by Gender and Ethnicity: Grade 3 Reading

	N	α	Raw Score		Scaled Score	
			Mean	SD	Mean	SD
Total	1284	0.89	30.15	9.47	401.23	29.62
Female	645	0.89	30.93	9.40	403.75	29.56
Male	623	0.89	29.36	9.52	398.67	29.58
American Indian	3	NA	NA	NA	NA	NA
Asian/Pacific Islander	22	0.91	34.00	9.59	414.09	30.71
Black/African American	435	0.89	27.09	9.24	391.46	28.01
Hispanic	50	0.89	27.10	10.08	391.84	30.50
White	563	0.88	33.28	8.67	411.15	27.92
Multi-Ethnic	81	0.89	30.51	9.17	402.69	28.98
Other	12	0.91	28.83	9.84	396.33	29.68

Note. NA indicates N < 10.

APPENDIX L

Operational Public School Frequency Distributions by Gender and for Total Population

Table L1. Operational Public School Frequency Distributions by Gender and for Total Population: Grade 3 Reading

		Female		Male		Total	
Raw Score	Scaled Score	Count	%	Count	%	Count	%
0	252	0	0.00	1	0.00	2	0.00
1	271	0	0.00	2	0.00	2	0.00
2	289	3	0.00	16	0.02	19	0.01
3	301	10	0.02	20	0.03	30	0.02
4	309	17	0.03	29	0.04	48	0.04
5	316	38	0.06	68	0.10	107	0.08
6	322	67	0.11	113	0.17	184	0.14
7	327	113	0.18	197	0.30	317	0.24
8	332	174	0.27	310	0.47	487	0.37
9	336	248	0.39	467	0.71	720	0.55
10	340	345	0.54	543	0.82	896	0.69
11	344	438	0.69	639	0.97	1088	0.84
12	347	491	0.77	733	1.11	1235	0.95
13	350	575	0.91	802	1.21	1386	1.07
14	354	601	0.95	907	1.37	1520	1.17
15	357	775	1.22	939	1.42	1725	1.33
16	360	756	1.19	1016	1.54	1786	1.37
17	363	798	1.26	1035	1.57	1847	1.42
18	366	854	1.35	1148	1.74	2013	1.55
19	369	1016	1.60	1266	1.92	2294	1.76
20	371	1027	1.62	1348	2.04	2394	1.84
21	374	1094	1.72	1362	2.06	2468	1.90
22	377	1201	1.89	1463	2.22	2675	2.06
23	380	1309	2.06	1439	2.18	2759	2.12
24	382	1256	1.98	1537	2.33	2806	2.16
25	385	1429	2.25	1611	2.44	3053	2.35
26	388	1484	2.34	1643	2.49	3147	2.42
27	390	1582	2.49	1732	2.62	3328	2.56
28	393	1606	2.53	1785	2.70	3407	2.62
29	396	1800	2.84	1875	2.84	3696	2.84
30	398	1770	2.79	2013	3.05	3799	2.92
31	401	2082	3.28	1980	3.00	4081	3.14
32	404	2144	3.38	2159	3.27	4316	3.32

		Female		Male		Total	
Raw Score	**Scaled Score**	**Count**	**%**	**Count**	**%**	**Count**	**%**
33	407	2346	3.70	2219	3.36	4585	3.53
34	410	2461	3.88	2369	3.59	4844	3.73
35	415	2603	4.10	2525	3.82	5143	3.96
36	416	2722	4.29	2702	4.09	5446	4.19
37	420	2921	4.60	2829	4.28	5767	4.44
38	423	2966	4.67	2847	4.31	5829	4.48
39	427	3109	4.90	3059	4.63	6185	4.76
40	432	3154	4.97	2856	4.33	6027	4.64
41	436	3021	4.76	2813	4.26	5851	4.50
42	441	2889	4.55	2563	3.88	5465	4.20
43	446	2615	4.12	2370	3.59	4992	3.84
44	452	2090	3.29	1870	2.83	3966	3.05
45	459	1582	2.49	1289	1.95	2876	2.21
46	468	1001	1.58	868	1.31	1869	1.44
47	480	548	0.86	431	0.65	979	0.75
48	499	262	0.41	186	0.28	448	0.34
49	518	55	0.09	38	0.06	93	0.07

Note. The sum of Females and Males is not equal to the total number of public school students due to missing values in gender.

APPENDIX M

Operational Public School Frequency Distributions by Ethnicity

Table M1. Operational Public School Frequency Distributions by Ethnicity – Grade 3 Reading

		American Indian		Asian/ Pacific Islander		Black/ African American		Hispanic		White		Multi-Ethnic		Other		Total	
RS	SS	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0	252	0	0.00	0	0.00	1	0.01	0	0.00	0	0.00	0	0.00	0	0.00	2	0.00
1	271	0	0.00	0	0.00	2	0.01	0	0.00	0	0.00	0	0.00	0	0.00	2	0.00
2	289	0	0.00	1	0.04	7	0.04	1	0.03	9	0.01	0	0.00	0	0.00	19	0.01
3	301	1	0.57	0	0.00	19	0.10	1	0.03	6	0.01	3	0.05	0	0.00	30	0.02
4	309	0	0.00	1	0.04	19	0.10	2	0.05	19	0.02	2	0.04	2	0.16	48	0.04
5	316	0	0.00	0	0.00	47	0.24	8	0.21	46	0.05	3	0.05	1	0.08	107	0.08
6	322	2	1.14	2	0.09	91	0.46	15	0.40	54	0.06	9	0.16	3	0.24	184	0.14
7	327	2	1.14	3	0.13	136	0.68	15	0.40	131	0.14	13	0.23	6	0.48	317	0.24
8	332	4	2.29	5	0.22	208	1.05	27	0.72	197	0.21	21	0.37	14	1.11	487	0.37
9	336	3	1.71	5	0.22	290	1.46	49	1.31	331	0.35	16	0.29	12	0.95	720	0.55
10	340	2	1.14	7	0.31	357	1.80	53	1.42	407	0.43	46	0.82	6	0.48	896	0.69
11	344	2	1.14	4	0.18	425	2.14	63	1.68	476	0.50	63	1.12	17	1.35	1088	0.84
12	347	4	2.29	7	0.31	452	2.27	76	2.03	603	0.63	49	0.87	18	1.43	1235	0.95
13	350	1	0.57	13	0.57	485	2.44	87	2.33	690	0.73	71	1.27	11	0.87	1386	1.07
14	354	2	1.14	9	0.40	513	2.58	76	2.03	802	0.84	74	1.32	16	1.27	1520	1.17
15	357	1	0.57	12	0.53	524	2.64	99	2.65	958	1.01	75	1.34	19	1.51	1725	1.33
16	360	4	2.29	17	0.75	562	2.83	84	2.25	988	1.04	68	1.21	24	1.90	1786	1.37
17	363	1	0.57	10	0.44	517	2.60	81	2.17	1078	1.13	101	1.80	23	1.82	1847	1.42
18	366	2	1.14	17	0.75	520	2.62	113	3.02	1188	1.25	103	1.84	27	2.14	2013	1.55
19	369	4	2.29	25	1.11	638	3.21	101	2.70	1355	1.43	106	1.89	24	1.90	2294	1.76
20	371	6	3.43	22	0.97	612	3.08	106	2.83	1450	1.53	118	2.11	27	2.14	2394	1.84
21	374	1	0.57	27	1.19	597	3.00	112	2.99	1512	1.59	130	2.32	34	2.70	2468	1.90
22	377	5	2.86	34	1.50	638	3.21	104	2.78	1699	1.79	116	2.07	29	2.30	2675	2.06
23	380	5	2.86	27	1.19	629	3.16	106	2.83	1778	1.87	138	2.46	34	2.70	2759	2.12
24	382	2	1.14	29	1.28	578	2.91	130	3.48	1844	1.94	141	2.52	32	2.54	2806	2.16
25	385	4	2.29	48	2.12	599	3.01	114	3.05	2053	2.16	134	2.39	44	3.49	3053	2.35
26	388	3	1.71	39	1.72	613	3.08	119	3.18	2147	2.26	137	2.44	34	2.70	3147	2.42
27	390	9	5.14	38	1.68	639	3.21	116	3.10	2260	2.38	174	3.10	34	2.70	3328	2.56
28	393	2	1.14	35	1.55	617	3.10	100	2.67	2378	2.50	173	3.09	38	3.01	3407	2.62
29	396	5	2.86	47	2.08	613	3.08	129	3.45	2641	2.78	152	2.71	39	3.09	3696	2.84
30	398	5	2.86	45	1.99	657	3.30	123	3.29	2685	2.83	177	3.16	32	2.54	3799	2.92

RS	SS	American Indian		Asian/ Pacific Islander		Black/ African American		Hispanic		White		Multi-Ethnic		Other		Total	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
31	401	9	5.14	46	2.03	676	3.40	103	2.75	2958	3.11	168	3.00	39	3.09	4081	3.14
32	404	7	4.00	51	2.26	670	3.37	117	3.13	3155	3.32	201	3.59	49	3.89	4316	3.32
33	407	4	2.29	68	3.01	622	3.13	131	3.50	3441	3.62	201	3.59	42	3.33	4585	3.53
34	410	8	4.57	82	3.63	625	3.14	110	2.94	3681	3.88	227	4.05	43	3.41	4844	3.73
35	415	6	3.43	70	3.10	633	3.18	110	2.94	3985	4.20	220	3.93	45	3.57	5143	3.96
36	416	6	3.43	97	4.29	569	2.86	120	3.21	4243	4.47	259	4.62	63	5.00	5446	4.19
37	420	10	5.71	102	4.51	622	3.13	124	3.31	4550	4.79	236	4.21	36	2.85	5767	4.44
38	423	3	1.71	92	4.07	554	2.79	119	3.18	4679	4.93	254	4.53	40	3.17	5829	4.48
39	427	6	3.43	127	5.62	537	2.70	117	3.13	4990	5.25	272	4.85	52	4.12	6185	4.76
40	432	8	4.57	129	5.71	460	2.31	115	3.07	4967	5.23	224	4.00	41	3.25	6027	4.64
41	436	4	2.29	138	6.10	368	1.85	87	2.33	4922	5.18	220	3.93	49	3.89	5851	4.50
42	441	4	2.29	164	7.25	297	1.49	85	2.27	4607	4.85	198	3.53	52	4.12	5465	4.20
43	446	8	4.57	125	5.53	272	1.37	83	2.22	4243	4.47	159	2.84	35	2.78	4992	3.84
44	452	1	0.57	139	6.15	168	0.85	39	1.04	3411	3.59	144	2.57	24	1.90	3966	3.05
45	459	5	2.86	104	4.60	113	0.57	34	0.91	2468	2.60	100	1.78	21	1.67	2876	2.21
46	468	0	0.00	92	4.07	55	0.28	23	0.61	1593	1.68	70	1.25	19	1.51	1869	1.44
47	480	4	2.29	71	3.14	25	0.13	10	0.27	831	0.87	26	0.46	8	0.63	979	0.75
48	499	0	0.00	30	1.33	9	0.05	3	0.08	393	0.41	8	0.14	2	0.16	448	0.34
49	518	0	0.00	5	0.22	1	0.01	1	0.03	81	0.09	4	0.07	1	0.08	93	0.07

Note. Missing values are coded as Other.

APPENDIX N

Operational Non-Public School Frequency Distribution

Table N1. Operational Non-Public School Frequency Distribution – Grade 3 Reading

Raw Score	Scaled Score	Total	
		Count	%
0	252	0	0.00
1	271	0	0.00
2	289	0	0.00
3	301	1	0.08
4	309	0	0.00
5	316	1	0.08
6	322	0	0.00
7	327	3	0.23
8	332	6	0.47
9	336	9	0.70
10	340	7	0.55
11	344	16	1.25
12	347	18	1.40
13	350	19	1.48
14	354	20	1.56
15	357	16	1.25
16	360	17	1.32
17	363	23	1.79
18	366	29	2.26
19	369	16	1.25
20	371	36	2.80
21	374	38	2.96
22	377	38	2.96
23	380	39	3.04
24	382	18	1.40
25	385	28	2.18
26	388	30	2.34
27	390	29	2.26
28	393	41	3.19
29	396	39	3.04
30	398	46	3.58
31	401	50	3.89

Raw Score	Scaled Score	Total	
		Count	%
32	404	55	4.28
33	407	43	3.35
34	410	40	3.12
35	415	61	4.75
36	416	59	4.60
37	420	55	4.28
38	423	52	4.05
39	427	59	4.60
40	432	31	2.41
41	436	57	4.44
42	441	57	4.44
43	446	37	2.88
44	452	23	1.79
45	459	11	0.86
46	468	9	0.70
47	480	2	0.16
48	499	0	0.00
49	518	0	0.00

APPENDIX O

Operational Interrater Analysis: G-Study

Table O1. Mean Variance Components and Interrater Reliability/Generalizability Coefficients for the One Facet Design and for the Two Facet Design Maximal Fully Crossed Subsets

	Subset Type	var(p)	var(i)	var(r)	var(pi)	var(pr)	var(ir)	var(pir)	rho2
G3R	Single	0.125	NA	0.000	NA	0.006	NA	NA	0.945
	Pair	0.045	0.012	0.000	0.081	0.000	0.000	0.006	0.948
	Triplet	0.046	0.012	0.000	0.079	0.000	0.000	0.006	0.948
	Quartet	0.042	0.012	0.000	0.081	0.000	0.000	0.007	0.944

APPENDIX P

Operational Subscale Reliability and Passing Bands

Table P1. Subscale Reliabilities and Passing Bands – Grade 3 Reading

	Raw Score Bands				
Content Standard	**Below**	**At**	**Above**	**Possible Score**	**Alpha**
Vocabulary	0 – 6	7 – 8	9 – 10	10	0.70
Reading process	0 – 6	7 – 10	11 – 16	16	0.75
Informational text	0 – 5	6 – 8	9 – 13	13	0.70
Literary text	0 – 4	5 – 7	8 – 10	10	0.62

APPENDIX Q

Operational Public School Frequency Distributions for Subscales

Table Q1. Operational Frequency Distribution for Subscales: Grade 3 Reading

	AV		IT		LT		RP	
Raw Score	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**	**Freq**	**Percent**
0	111	0.09	348	0.27	811	0.62	533	0.41
1	385	0.30	1542	1.19	2760	2.12	1886	1.45
2	1207	0.93	3394	2.61	5819	4.48	3491	2.69
3	2472	1.90	5762	4.43	8921	6.86	4956	3.81
4	4202	3.23	7730	5.95	11403	8.77	5989	4.61
5	6093	4.69	9522	7.32	13857	10.66	6857	5.27
6	9078	6.98	11833	9.10	16984	13.06	7511	5.78
7	13830	10.64	14009	10.78	20261	15.59	8187	6.30
8	24165	18.59	16522	12.71	22054	16.96	9014	6.93
9	46382	35.68	18168	13.98	17624	13.56	10061	7.74
10	22075	16.98	17741	13.65	9506	7.31	11072	8.52
11			13669	10.51			12085	9.30
12			7643	5.88			12937	9.95
13			2117	1.63			12921	9.94
14							11558	8.89
15							7747	5.96
16							3195	2.46

Note. AV: Vocabulary; IT: Information Text; LT: Literary Text; RP: Reading Process.

APPENDIX R

Subscale Intercorrelations

Table R1. Subscale Intercorrelations – Grade 3 Reading

		Acquisition of Vocabulary	Informational Text	Literary Text
Grade 3	Informational Text	0.67		
	Literary Text	0.63	0.68	
	Reading Process	0.69	0.76	0.73