DATA RECOGNITION
DRC
CORPORATION

Beyond the Numbers

A Guide to Interpreting and Using the Results of Standardized Achievement Tests

TERRANOVA 3®

Beyond the Numbers

A Guide to Interpreting and Using the Results of Standardized Achievement Tests

DATA RECOGNITION
DRC
CORPORATION

Developed and published by Data Recognition Corporation, Copyright© 2017 by Data Recognition Corporation. Only authorized customers may copy or print the document. Any other use or reproduction of this document, in whole or in part, requires written permission of the publisher.

The Value of Standardized Achievement Tests

Achievement tests are designed to measure the things that a student knows and can do.

A well-planned, comprehensive assessment program built upon a quality test provides valuable information about students that can support teaching in many areas, including:

- Evaluation of Student Progress
- Needs Assessment
- Instructional Program Planning
- Curriculum Analysis
- Program Evaluation
- Class Grouping
- Administrative Planning and Direction

A sound assessment and evaluation program reinforces educational decisions.

It can positively impact individual students as well as parents or guardians, groups or classes of students, schools, and communities.

Test data can be used to:

- Aid teachers and administrators in improving instructional plans
- Keep parents informed about their children's progress
- Inform the community of the general progress being made by a school or school system

Appropriate Uses of Test Data

To Help Teachers Teach

Test results are most valuable when the testing information is used to support instruction in the classroom. Test results can also be used to help make critical administrative decisions that affect students and their families, such as eligibility for placement into special programs. Additionally, test results can be used to evaluate the effectiveness of the curriculum and the instructional program and to keep the public informed of progress being made by schools.

Information You Can Trust

Such critical uses of test results place a heavy responsibility on the test publisher, who must ensure that the results are accurate and that clear, reliable information about the interpretation of results is provided. Data Recognition Corporation welcomes this responsibility and works to produce quality tests that provide the most valid and reliable information possible.

How to Use This Brochure

This publication is designed to help you use the information provided by achievement tests wisely and appropriately.

It is meant to supplement, not take the place of, the information found in such publications as examiner's manuals, norms books, test coordinator's handbooks, teacher's guides, and technical reports. DRC's Assessment Solutions Consultants may also assist school staff in making the best use of the information provided by DRC's tests. You may contact your Assessment Solutions Consultant by calling 800.538.9547 option 2. You may also find helpful information about DRC's tests at our website www.datarecognitioncorp.com

Translating Standards into Learning

No matter how education systems change and grow, there will be a need for quality assessments that present reliable and accurate information about student achievement. DRC is committed to filling this need.

DRC offers a number of innovative reporting options and delivery systems that help states, districts, and schools meet the requirements of the NCLB. These include:

- Individual and group performance reports on all assessments;
- Custom reports based on state standards;
- Online test reports;
- Reports delivered on CD-ROM.

DRC's test reports and customizable reporting systems help translate test results into useful, easy to understand information that supports and enhances teaching strategies for teachers and parents. All reports are designed to present complex achievement data simply in easily understood formats and terms.

Table of Contents

General Testing Guidelines 4

Types of Information Provided by Achievement Tests 5

- Norm-Referenced Information 6
- Norm-Referenced Scores Reported for Individual Students 6
- Interpreting Norm-Referenced Scores for Individuals 10
- Norm-Referenced Scores Reported for Groups 12
- Interpreting Norm-Referenced Scores for Groups 12
- Anticipated Achievement Scores 13
- Criterion-Referenced Information 14
- Interpreting Criterion-Referenced Information 15
- Performance Level Information 16
- Interpreting Performance Level Information 17

Glossary of Assessment Terms 18

Additional Publications That Provide Useful Information 20

Concordance between the National Percentile and the Mean Normal Curve Equivalent 21

General Testing Guidelines

Certain principles and cautions should be kept in mind when interpreting and using the results of any group-administered standardized achievement test.

Consider Standard Error of Measurement

DRC develops the most accurate, valid, and reliable assessments available; however, all test scores should be interpreted with Standard Error of Measurement in mind.

(See "Standard Error of Measurement" page 19.)

Use All Available Data

When making critical decisions that affect the life of a student, all available information should be considered, including previous test scores, class grades, and teacher judgment.

Don't Hesitate to Question Results

If there is good reason to question the accuracy of a test result, that test result should not be made part of a student's permanent record. For example, if the student was ill or unduly stressed while taking the test, the test results may not accurately reflect the student's knowledge or ability.

Take a Broad Perspective

Any test result that is at great variance with other information about the student, such as previous test scores and class grades, should be interpreted with caution.

Interpret Chance Scores with Caution

Scores in the "chance range" of the test-that is, scores that could be obtained by random guessing-should be interpreted cautiously. Such scores often mean that the level administered was too difficult. Retesting with a lower level of the test would probably provide more useful information about the student's specific needs. Locator tests are available to help you determine the proper level to administer to students in question.

Interpret Perfect Scores with Caution

When all items are answered correctly, test results should be regarded as an uncertain description of achievement. Retesting with a higher level of the test would provide better information about the student's true top performance level. Locator tests are available to help you determine the proper level to administer to students in question.

Administer as Directed

Norm-referenced scores are valid only when the test is administered following the same directions and time limits under which the test was standardized. For *TerraNova,™ Third Edition*, this includes administering the test with appropriate accommodations. When students are unable to take the test under standard conditions, administering the test with modified directions and time limits may yield valuable information for diagnostic purposes, but it may impact the norm-referenced scores.

Consider the Size of the Test Group

Average (mean and median) scores based on small numbers of students--generally fewer than 25-should be interpreted with caution. One or two extreme scores can greatly affect the mean and median in such instances.

Test Results are Dated

Note that the time-bound nature of national norms should be recognized. The norms for a standardized achievement test describe the levels of achievement that prevailed at the time the test was standardized.

Keeping these principles and cautions in mind will help ensure that test information is used in ways that best serve the interests of the student, the school, and the community.

Types of Information Provided by Achievement Tests

Most educators today want an achievement test that supports and informs instruction and that monitors changes in student performance. They want a test that provides three basic kinds of information:

- Accurate norm-referenced information
- Relevant criterion-referenced information
- Standards-based performance information

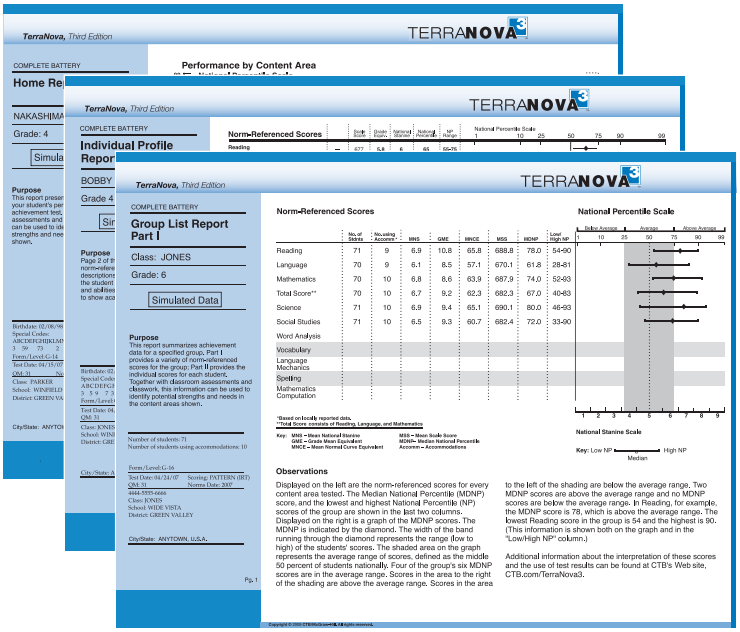
This section discusses each kind of information and defines the scores that convey that information.

Sample Score Reports

Throughout this publication, you will find sample score report data and explanations to help interpret that data. Samples of actual score reports are available from DRC Assessment Solutions Consultants and Customer Care. DRC is recognized for the readability and usefulness of our score reports. DRC assessments provide relevant, accurate, and useful norm-referenced, criterion-referenced, and standards-based performance level information. Score reports offer further insight and greater understanding about test information.

TerraNova, Third Edition

DRC's newest assessment series, *TerraNova, Third Edition*, is an excellent choice for schools planning a testing program designed to meet current and future assessment and reporting needs. This assessment links innovation and quality through its design, content, psychometric thoroughness, and graphic reporting. To learn more about DRC's full range of standardized achievement tests, please contact your DRC Assessment Solutions Consultant at 800.538.9547 option 2, or visit www.datarecognitioncorp.com.



TERRANOVA 3®

Norm-Referenced Information

Norm-referenced test information permits the achievement of individuals and groups of students to be compared with national performance.

A variety of scores can be used to make these comparisons. Brief, non-technical descriptions of each type of score follow. The first two sections provide an overview of scores reported for individual students and how to interpret them. The second two sections present scores reported for groups of students.

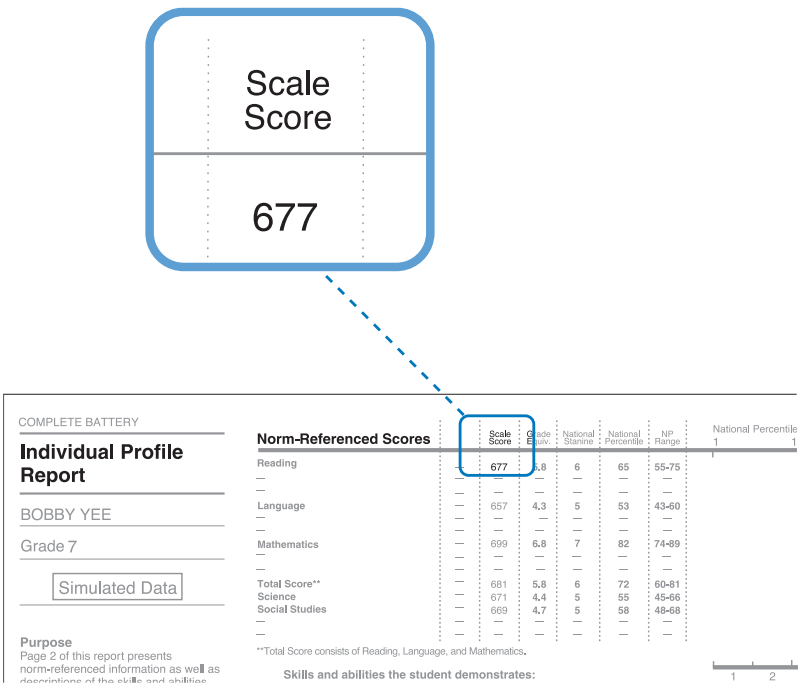
Norm-Referenced Scores Reported for Individual Students

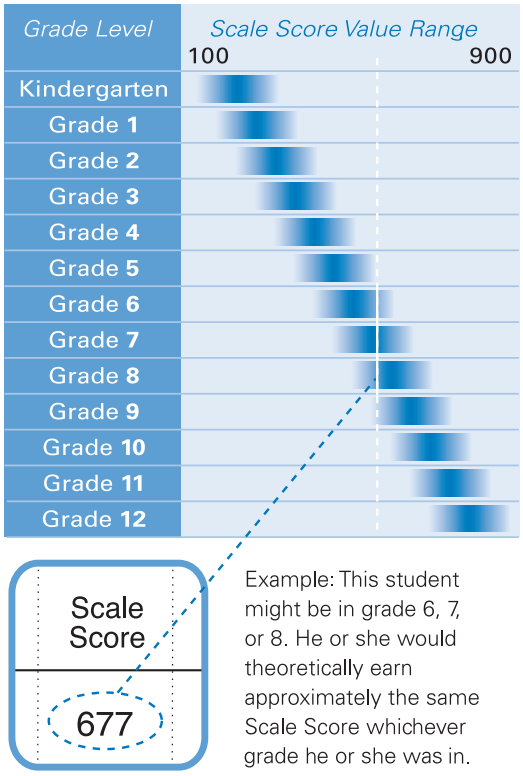
Scale Score

DRC offers the Item Respose Theory (IRT) method of scoring to derive a Scale Score. The Scale Score (SS) is the basis for norm-referenced scores. The Scale Score describes achievement on a continuum that in most cases spans the complete range of Kindergarten through Grade 12. These scores can range in value from approximately 100 to 915.

For a test such as *TerraNova, Third Edition*, Scale Scores are "level-independent." The overlapping levels are linked to each other in such a way that, when *TerraNova, Third Edition* is administered to a group of students, any given student is expected to obtain the same Scale Score regardless of the form, level, or scoring method used, within the limits of the Standard Error of Measurement (SEM) and the floor and ceiling of the test level used. Scale Scores have the advantage of being an "equal-interval" scale that can be treated arithmetically. For example, they can be added and divided for purposes of averaging.

A limitation of Scale Scores is that they have no inherent or readily apparent meaning for non-experts. They are used most often for statistical analyses and research purposes in deriving other scores, such as the Performance Level, Grade Equivalent, National Percentile, National Stanine, and Normal Curve Equivalent, each of which is described in this booklet. (A more technical explanation of the derivation of the Scale Score is found in the *TerraNova, Third Edition* Technical Bulletin.)





Example: This student might be in grade 6, 7, or 8. He or she would theoretically earn approximately the same Scale Score whichever grade he or she was in.

Scale Scores and Item Response Theory

DRC offers the Item Respose Theory (IRT) of calculating Scale Scores. The IRT pattern scoring method takes into account which items the student got correct and produces better test information, less measurement error, and greater reliability than number-correct scoring. We provide IRT pattern scoring because it produces more accurate scores for individual students.

DRC's Assessment Solutions Consultants can provide additional information about the application of IRT in the development of DRC tests. No other major publisher has made the investment required to offer the many advantages of the more complete use of information that IRT provides.

Grade Equivalent

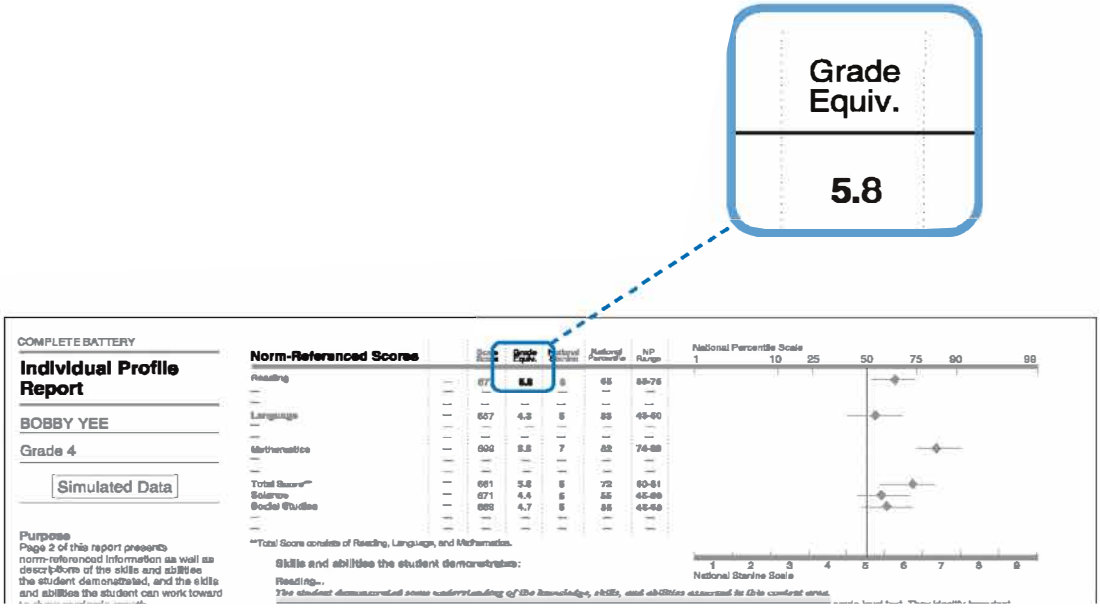
A Grade Equivalent (GE) is useful in describing comparative levels of achievement in the elementary grades when the skills measured by an achievement test (such as *TerraNova, Third Edition*) are generally taught and when there are fairly regular gains in skill development.

A Grade Equivalent indicates the year and month of school for which a student's level of performance is typical. A GE of 6.2, for example, is interpreted to mean that the student in question is achieving at a level that is typical of students who had completed the second month of Grade 6 at the time the test was standardized (September being designated as .0, June as .9).

What GEs Mean

Grade Equivalents that are within approximately two years of the student's actual grade placement are generally accurate descriptions of the student's level of achievement, since adjacent levels of *TerraNova* were directly linked during standardization. Extreme GEs—generally those more than two years above or below the student's actual grade level—are based on statistical links between several test levels and must be interpreted with caution. A GE of 6.6 obtained by a student in Grade 3, for example, does not mean that the student is capable of doing sixth-grade work. It means only that this student is scoring well above average for Grade 3. In most cases, intervening instruction must take place if the student is to be prepared to succeed when he or she reaches Grade 6.

The GE is not an equal-interval scale—that is, the distance between two GEs has different meaning at different points on the scale. GEs should not be treated arithmetically; they should not be averaged directly or added or subtracted from one another.



Norm-Referenced Information Reported for Individual Students

National Percentile

The National Percentile (NP) is the best score for describing achievement test results to persons outside the test and measurement community. A National Percentile score (or rank) of 65, for example, means that the student's score is higher than the scores of 65 percent of the students in the norming group for that student's grade at that time of year. Percentiles are essentially "counting scores" that designate the percentage of students in the norming sample for a given grade whose scores fell below a certain point.

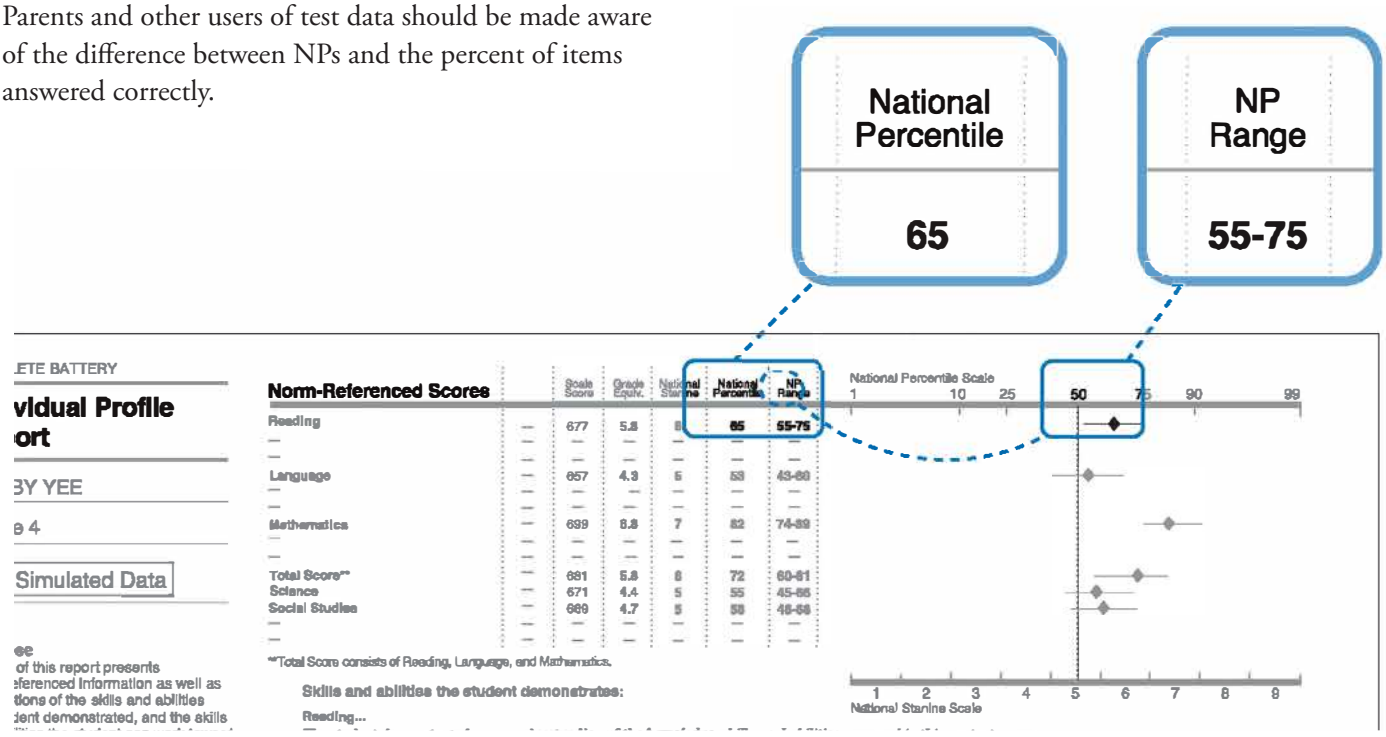
Understanding the NP

The 25th NP, then, is the score below which 25 percent of the norm group scored. The 50th NP is the middle score—also called the Median—above and below which half of the students scored. The 75th NP is the score below which 75 percent of the students in the norming sample scored; conversely, only 25 percent had scores higher than that NP.

The NP is a non-equal-interval scale, meaning that the distance between two NPs has different meaning at different points on the scale. NPs should not be treated arithmetically; that is, they should not be averaged directly. The NP can also be confused with the percent of items answered correctly. Parents and other users of test data should be made aware of the difference between NPs and the percent of items answered correctly.

As noted previously, DRC uses Scale Scores for calculation of National Percentiles. There are three Norms Books for *TerraNova*, *Third Edition* containing tables for each grade for converting Scale Scores to National Percentiles (as well as to other derived scores, explained below). There is one book for the fall trimester (September through November), one for the winter trimester (December through February), and one for the spring trimester (March through June).

For even greater precision and for ready compliance with federal guidelines for the evaluation of Title I programs, DRC bases National Percentiles and Normal Curve Equivalents on the specific quarter month during which an assessment is administered. Empirical norms are used during the authorized five-week "windows," fall and spring, and interpolated norms are used outside those windows. This service is provided automatically and at no additional cost.



National Stanine

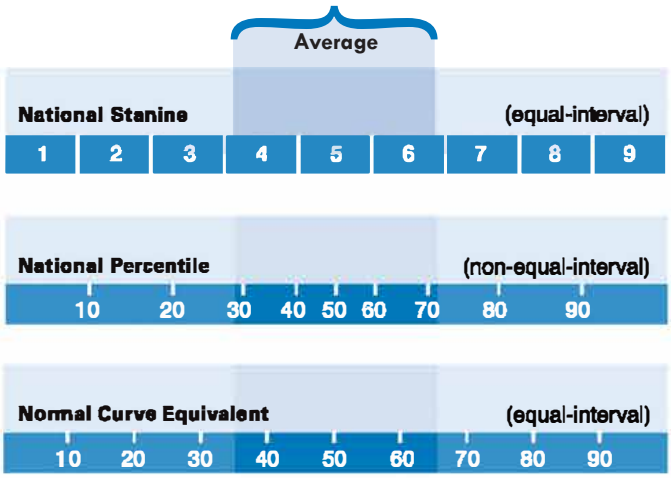
The National Stanine is a scale that divides the scores of the norming population into nine groups. This score has the advantage of being an equal-interval scale, meaning that it, unlike the National Percentile and the Grade Equivalent, can be treated arithmetically.

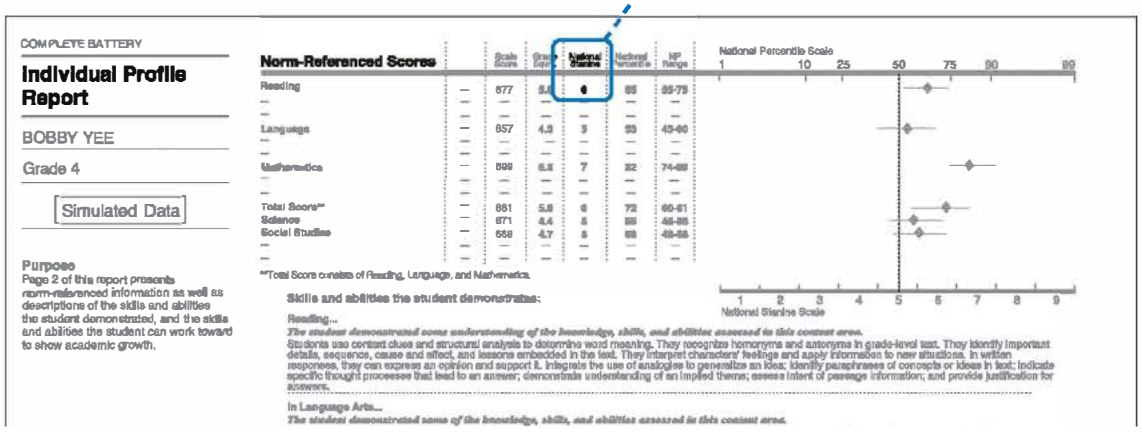
Because Stanines are single-digit numbers, they are less likely than Percentiles to be confused with the percentage of items answered correctly. Their major disadvantage is their lack of precision. For example, a student with a National Stanine score of 5 could have a National Percentile as low as 41 and as high as 59.

Normal Curve Equivalent

The Normal Curve Equivalent (NCE) was developed specifically for use in the evaluation of Title I programs. It is used to aggregate "Gain Scores" from different programs, even those using different standardized tests. It is generally not recommended for use in reporting scores of individual students, since it is easily confused with National Percentiles and has no inherent or readily apparent meaning.

The NCE, unlike the NP, is an equal-interval scale and can be treated arithmetically. There is a constant relationship between National Stanines, National Percentiles, and Normal Curve Equivalents. A given National Percentile will always correspond to a given Normal Curve Equivalent. NPs and NCEs, however, are coincident only at the 1st, 50th, and 99th points.





Norm-Referenced Information Reported for Individual Students

Interpreting Norm-Referenced Scores for Individuals

The Individual Profile Report provides norm-referenced information (and criterion-referenced information) for individual students. This page explains norm-referenced scores for the example shown below for a Grade 7 student.

1. Scale Score (SS)

The Scale Score in Reading is 677. Since Scale Scores have little inherent meaning, understanding the Reading performance requires looking at the student's other scores, which are based on the Scale Score.

2. Grade Equivalent (GE)

Based on the SS, the student's Grade Equivalent in Reading is 8.5. This means that the student's Reading achievement is equivalent to the average student who has completed the fifth month of Grade 8.

3. National Stanine (NS)

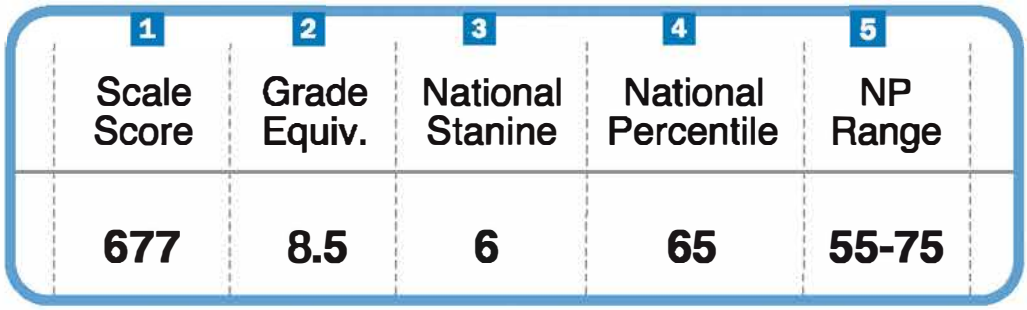
The student is placed in National Stanine 6. National Stanines 4, 5, and 6 are in the average range, thus this student's Reading achievement can be interpreted as average, or in the high end of the average range.

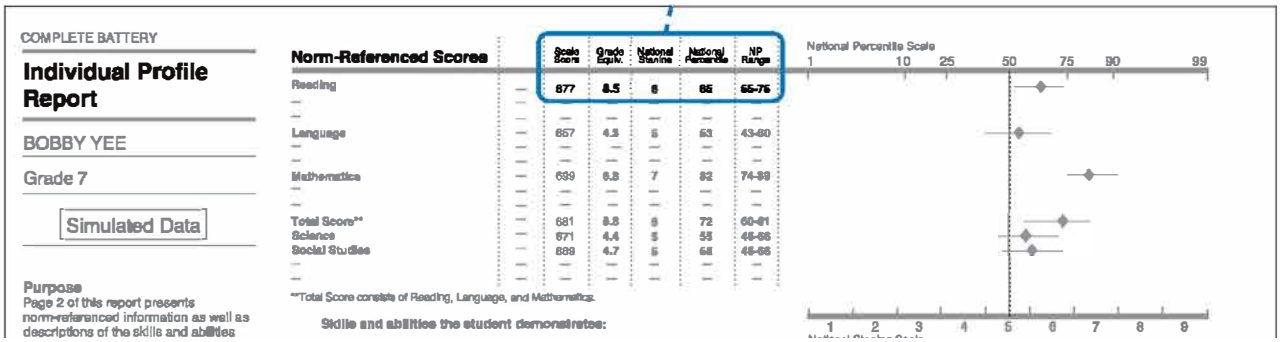
4. National Percentile (NP)

The SS places the student at the 65th NP, meaning the student's Reading achievement is better than that of 65% of students nationwide. Since National Percentiles of 25-75 are considered the average range, this student's achievement can be interpreted as average, or in the upper end of the average range.

5. National Percentile (NP) Range

The student's NP Range reflects the fact that all test results contain some measurement error, and that if the student had taken the Reading test numerous times, the NP would most likely be between 55 and 75. In other words, this student's "true score" on the Reading test is likely to be between the 55th and 75th percentiles.





Mean Scale Score (MSS)

The Mean Scale Score (MSS) is computed by adding all valid Scale Scores obtained by a group of students and dividing by the number of valid scores. It is the "arithmetic mean" or "arithmetic average." The MSS reported for the local group may be compared with that for the norm group.

Grade Mean Equivalent (GME)

Since GEs should not be averaged directly, the Grade Mean Equivalent (GME) is used to describe the mean, or "average," performance of a group of students in terms of Grade Equivalents. Derived from the Mean Scale Score, the GME describes the year and month of school at which the local group's MSS equals the National Mean. If a Mean Scale Score of 810, for example, converts to a GME of 6.6, it indicates that 810 is the Mean National Scale Score for students who have completed the sixth month of Grade 6.

The GME is the best description of the "average" Grade Equivalent for a group of students. A more detailed explanation of its rationale and derivation is found in the *TerraNova, Third Edition* Technical Bulletin.

Median National Percentile (MDNP)

Since National Percentiles are "counting scores," averages for NPs on DRC tests are reported in terms of the Median. The Median is the middle score in a set of ranked scores. It is sometimes called the "counting average" because it is actually determined by counting up into the distribution until the score that divides the distribution into two equal halves is reached.

If a group of students has a Median National Percentile (MDNP) of 5O, it is an "average" group. If the MDNP is higher than 5O, the group is "above average." An MDNP of less than 5O represents "below average" in relation to the sample of students upon which the test was standardized.

Medians, unlike Means, have the characteristic of being unaffected by extreme scores. Therefore, many testing specialists prefer the Median when comparing groups with non-symmetrical (i.e. many high scores or many low scores compared with scores spread evenly from an NP of 1 to an NP of 99) score distributions.

Mean Normal Curve Equivalent (MNCE)

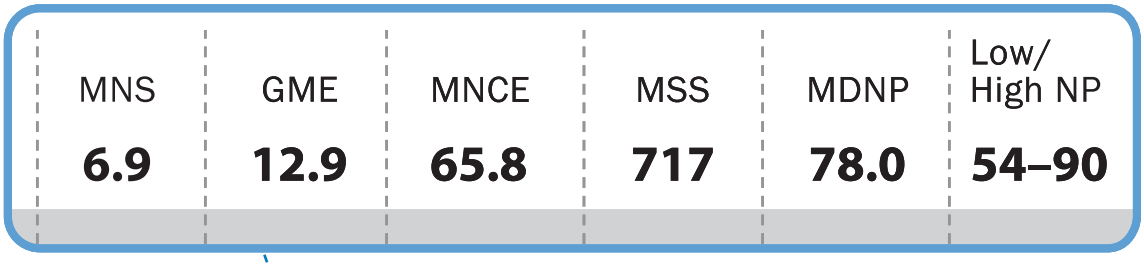
Since the NCE is an equal-interval scale, it can be treated arithmetically; therefore, the Mean NCE (MNCE) is computed by adding NCE scores of all students in a group with valid scores, then dividing by that number of students. It is the simple arithmetic average.

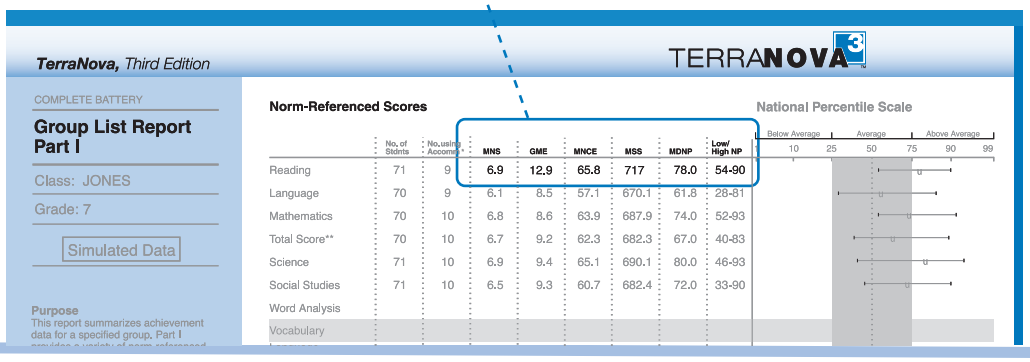
National Percentile of the Mean NCE (NP of the MNCE)

The NP of the MNCE is sometimes used for reporting group data. This score is simply the NP corresponding to the MNCE.

(See chart on page 21)

The widely used Title I Evaluation Model assumes that the MNCE will remain constant from one testing time to another, given normal growth in achievement. Stated another way, a "zero" MNCE gain score indicates normal growth in achievement from pre-test to post-test.





Interpreting Norm-Referenced Scores Reported for Groups

Interpreting Norm-Referenced Scores for Groups

The Group List Report, Part I provides norm-referenced scores for groups of students, such as a class. The explanations on this page refer to the example below to explain norm-referenced scores for a Grade 7 class of 71 students.

1. Mean National Stanine (MNS)

The MNS is the arithmetic average of the National Stanines for a group of students. Stanine scores of 4, 5, and 6 are defined as the average range, therefore, this group's MNS of 6.9 in Reading indicates that the mean score for this group is slightly above average.

2. Grade Mean Equivalent (GME)

The Reading Grade Mean Equivalent of 12.9 for the sample class indicates that the class's average Reading achievement is equivalent to that of an average student who has completed the ninth month of Grade 12.

3. Mean Normal Curve Equivalent (MNCE)

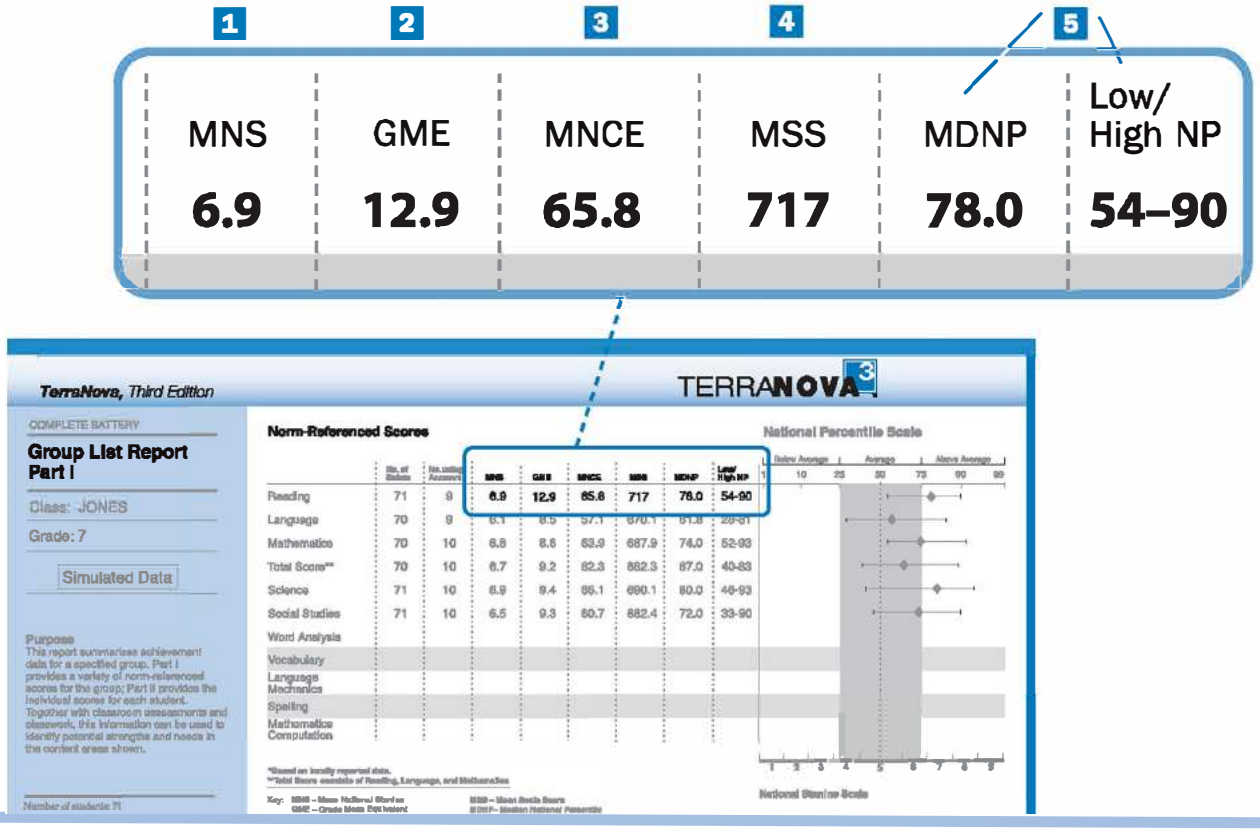
The MNCE of 65.8 in Reading is simply the arithmetic average of the Normal Curve Equivalent for the 71 students in the sample class. As shown in the graph on page 9, NCE scores of 36–64 are in the average range. The sample class's Reading MNCE of 65.8 indicates that the mean score for this group is slightly above average.

4. Mean Scale Score (MSS)

A Mean Scale Score, like a Scale Score, has little inherent meaning; therefore, understanding the class's Reading performance requires a review of the other norm-referenced scores provided.

5. Median National Percentile (MDNP)

The students in the sample class had Reading NPs ranging from a low of 54 to a high of 90 (Low/High NP). The Reading MDNP of 78.0 shows that, for this class, half of the student NP scores fell below 78 and half of the NP scores were higher than 78. The MDNP for Reading indicates that this group is slightly above average—the same result shown by the MNCE and the MNS.



Anticipated Achievement Scores

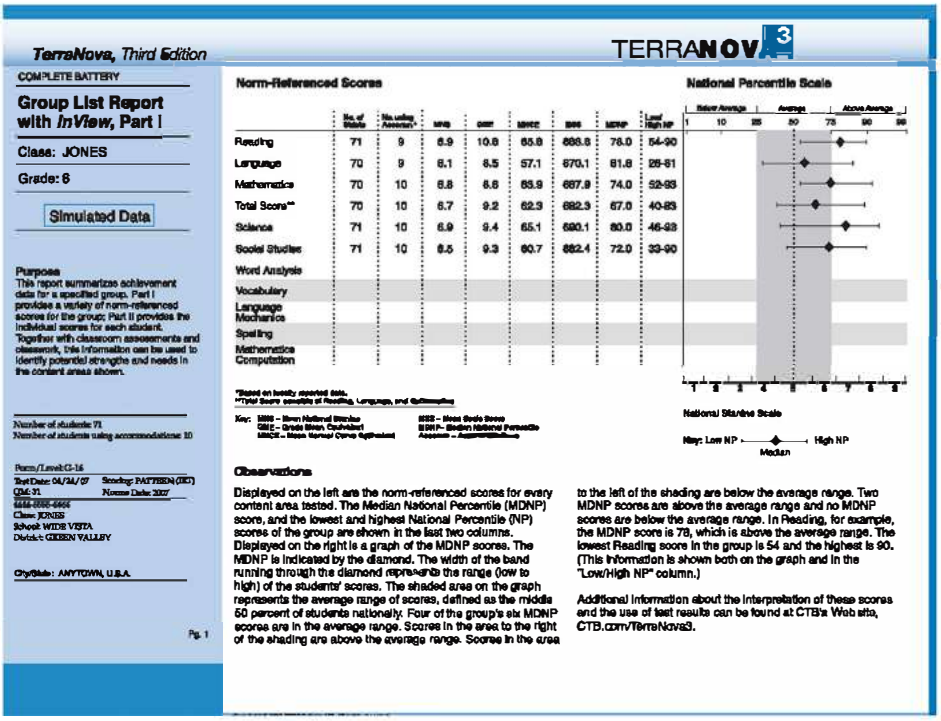
Schools that use *TerraNova* in combination with *InView*™ will see four types of "Anticipated Achievement" scores:

- Anticipated Achievement Scale Scores (AASS)
- Anticipated Achievement Grade Equivalents (AAGE)
- Anticipated Achievement National Percentiles (AANP)
- Anticipated Achievement Normal Curve Equivalents (AANCE)

Each of these scores represents an estimate of the average achievement score for students of similar age, grade, and cognitive abilities as measured by *InView*.
The multiple regression formulas used to compute Anticipated Achievement scores include as predictors age, grade, and scores on each of the subtests.

The difference between the student's actual score (also called the "Obtained Score") and the Anticipated Score is computed and, if the difference is statistically significant at the .20 level, a plus sign(+) or a minus sign(-) appears on the report beneath those test scores.
The "+" indicates that the Obtained Score is significantly higher than the Anticipated Score.
The "-" indicates that the Obtained Score is significantly lower than the Anticipated Score. Stated another way, the .80 level of confidence means that a difference is considered to be significant only if a difference that large would occur by chance alone 20 percent or less of the time.

Schools typically use the combination of an achievement test and a cognitive abilities test to identify underachieving students, or to demonstrate that they are helping students achieve at a higher level than expected.



TerraNova, *Third Edition* Group Level Report with *InView*, Part I.

Criterion-Referenced Information

The concept of reporting test information on the basis of students' levels of mastery of instructional content standards was pioneered by DRC in 1977 with the publication of the California Achievement Tests®, Forms C and D (CAT™ C and D). Since 1977, DRC has continued to be the industry leader in psychometric innovation and excellence.

The criterion-referenced information for *TerraNova, Third Edition* makes use of a combination of Bayesian Estimation and IRT procedures to produce the most accurate objective-mastery scores possible from an achievement test.

Objectives Performance Index

All versions of *TerraNova* yield criterion-referenced scores reported in terms of an Objectives Performance Index (OPI). An OPI is reported for each of the instructional objectives measured by the particular version of *TerraNova*. Each objective is measured by at least four items. DRC has established a standard of four items as the minimum number needed to produce reliable information regarding objective mastery. Other publishers continue to report objective-level information based on only 1 or 2 items–clearly not enough information for reliable decision making.

The OPI makes test results both understandable and useful for the teacher in planning effective learning strategies and activities. The OPI is an estimate of the number of items a student would be expected to answer correctly if there had been 100 similar items for that objective. For example, an OPI of 65 on a given objective means that if the student were given 100 similar items, the student would be expected to answer 65 of them correctly.

The OPI scale runs from "O" to "100."

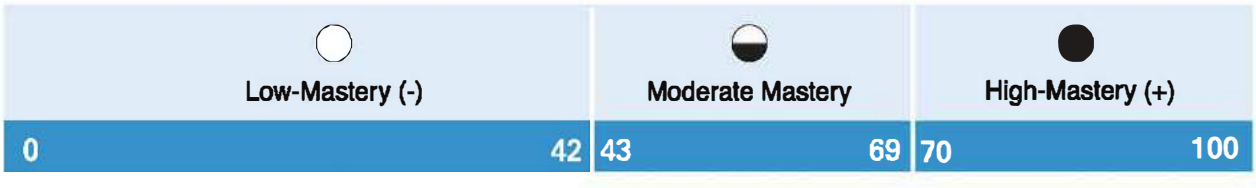
OPI Score Levels

For *TerraNova*, SUPERA®, OPI scores between O and 49 are regarded as the Low Mastery level. Scores between 50 and 74 are regarded as indications of Moderate Mastery. Scores of 75 and above are regarded as the High Mastery level.

For *TerraNova, Third Edition*, DRC used a more empirical approach to determine mastery levels on an objective-by-objective basis. DRC conducted a conference of teachers and content experts from across the country to develop the *TerraNova, Third Edition* Performance Levels cutscores. The mastery cutscores for *TerraNova, Third Edition*'s objectives were determined by:

- Computing the expected OPI for students at the "Proficient" (Level 4) cutscore and for the "Progressing" (Level 2) cutscore for each content area.
- Using these two expected OPIs, three regions on the OPI scale were defined for each objective on an objective-by-objective basis.

Mastery Levels for *TerraNova, Third Edition*



Criterion-Referenced Information

Using OPI Scores

The OPI can be used to help identify those objectives within a content area for which further instruction is indicated and those for which a satisfactory mastery level has already been attained. To assist in interpreting the meaning of differential mastery levels, the percentage of students in the norming sample who mastered each objective and the mean OPI are reported in technical publications such as the *TerraNova, Third Edition* Technical Bulletin as well as in several score reports, such as the Objectives Report. This information can help determine what is a reasonable expectation of mastery on a given objective.

Interpreting Criterion-Referenced Information

Interpreting Criterion-Referenced Information

The Objectives Summary provides summary criterion-referenced scores for groups of students. The explanations on this page refer to a sample school with three Grade 4 classes. Objectives Reports for groups of students that include individual student data are also available.

1. Percent of Students Mastering Each Objective

This report section shows the percent of students in the school, nation, and each class within the school that achieved high mastery for each objective (demonstrated the skills on that objective that teachers expected of a Proficient student). In the sample, 42% of Total School students mastered Basic Understanding of Reading, compared with 45% of students representing the national reference group, for a Difference of three percentage points. For the three classes that make up the Total School at this grade level, mastery ranged from 36% for Polk to 48% for Smith.

1.

Objectives Performance Index (OPI)

OPI is the estimated number of items correct out of 100

- ● Mastery (Range: 70 – 100 correct)
- ● Partial Mastery (Range: 43 – 69 correct)
- ○ Non- Mastery (Range: 0 – 42 correct)

** National Reference Group Grade 4.8

Percent of Students Mastering Each Objective

	Total School	National **	Difference	CLASS: OTTENWALTER	CLASS: POLK	CLASS: SMITH
Reading						
02 Basic Understanding	42	45	–3	42	36	48

TerraNova, Third Edition

COMPLETE BATTERY

Objectives Report, Part I (Part II is not shown)

Class: JONES

Grade: 4

Simulated Data

Purpose
This report summarizes achievement data for a specified group. Part I provides a variety of norm-referenced scores for the group; Part II provides the individual scores for each student. Together with classroom assessments and

Objectives Performance by School/Class

Objectives Performance Index (OPI)
OPI is the estimated number of items correct out of 100
● Mastery (Range: 70 – 100 correct)
● Partial Mastery (Range: 43 – 69 correct)
○ Non- Mastery (Range: 0 – 42 correct)
** National Reference Group Grade 4.8

Percent of Students Mastering Each Objective

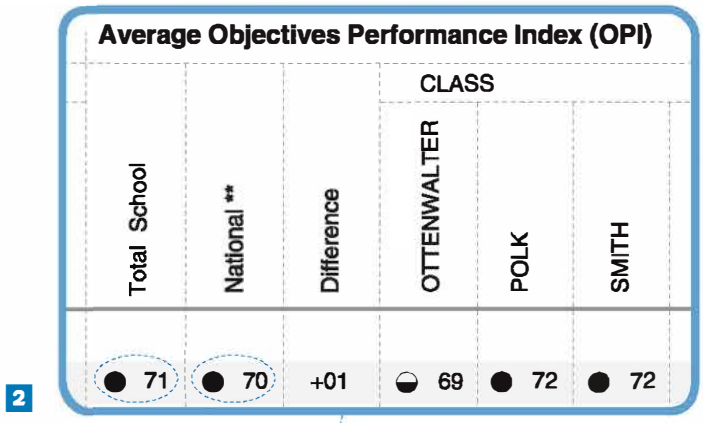
	Total School	National **	Difference	CLASS: OTTENWALTER	CLASS: POLK	CLASS: SMITH
Reading						
02 Basic Understanding	42	45	–3	42	36	48
03 Analyze Text	42	47	–6	38	36	52
04 Evaluate & Extend Meaning	27	30	–3	38	21	22
05 Identify Reading Strategies	17	16	+1	29	7	16
Language						
07 Sentence Structure	61	47	+14	63	57	63
08 Writing Strategies	57	39	+18	58	57	56
09 Editing Skills	43	32	+11	50	36	44
Mathematics						

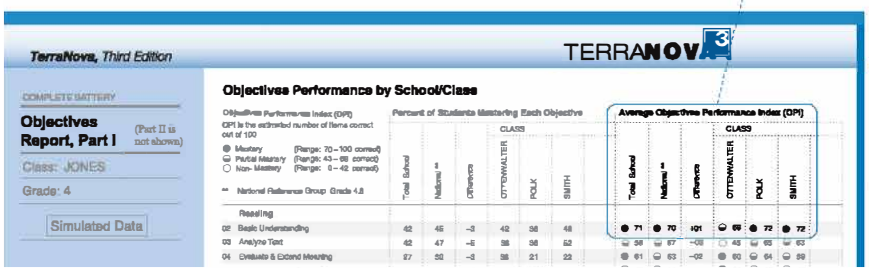
Average Objectives Performance Index (OPI)

	Total School	National **	Difference	CLASS: OTTENWALTER	CLASS: POLK	CLASS: SMITH
Reading						
02 Basic Understanding	● 71	● 70	+01	● 69	● 72	● 72
03 Analyze Text	● 68	● 67	+00	○ 45	● 65	● 83
04 Evaluate & Extend Meaning	● 61	● 63	–02	● 60	● 64	● 66
05 Identify Reading Strategies	○ 47	○ 44	+03	● 54	○ 42	○ 45
Language						
07 Sentence Structure	● 76	● 67	+09	● 78	● 75	● 73
08 Writing Strategies	● 72	● 63	+09	● 75	● 71	● 71
09 Editing Skills	● 69	● 61	+08	● 71	● 69	● 68

2. Average Objectives Performance Index (CPI)

The Average OPI of the students in the sample school's Grade 4 classes in Basic Understanding of Reading is 71, which is in the High Mastery category *(see page 14)*. The National Average OPI for this objective is 70, also indicating High Mastery. Two of the three classes within the sample school had Average OPIs for this objective in the High Mastery range, and teacher Ottenwalter's class had an average OPI in the Moderate Mastery Range, indicating that additional instruction is needed to reach the High Mastery range.





Performance Level Information

In addition to reporting student performance with a specified set of objectives (criterion-referenced scores), student performance can be reported in terms of five performance levels that describe a Pathway to Proficiency. DRC can also develop custom performance level reports that predict performance on state assessment, with 3–7 levels possible.

Each performance level represents standards of performance for each content area assessed (Reading, Language, Mathematics, Science, and Social Studies). Performance level scores provide a description of what students should be able to do in terms of the content and skills assessed by *TerraNova, Third Edition*.

Performance Level	*Proficiency Level*
Level 1	**"Starting Out"** *(Grades 1–2)* **"Step 1"** *(Grades 3–12)*
Level 2	**"Progressing"**
Level 3	**"Nearing Proficiency"**
Level 4	**"Proficient"**
Level 5	**"Advanced"**

Monitoring Performance Level Growth

Student progress is monitored both within a grade span and across grade spans, so that all students have an opportunity to demonstrate progress toward the desired goal of proficiency. The *TerraNova, Third Edition* performance level scores are reported for each of these grade spans:

- Grades 1–2 Primary
- Grades 3–5 Elementary
- Grades 6–8 Middle
- Grades 9–12 High

All students should work toward reaching proficiency by the time they complete the grade span. For example, in the elementary grade span, students should work toward reaching Level 4 or 5 by the end of Grade 5.

See DRC's Performance Level section of the *Teacher's Guide to TerraNova, Third Edition*, for more detailed information about performance level reports.

Interpreting Performance Level Information

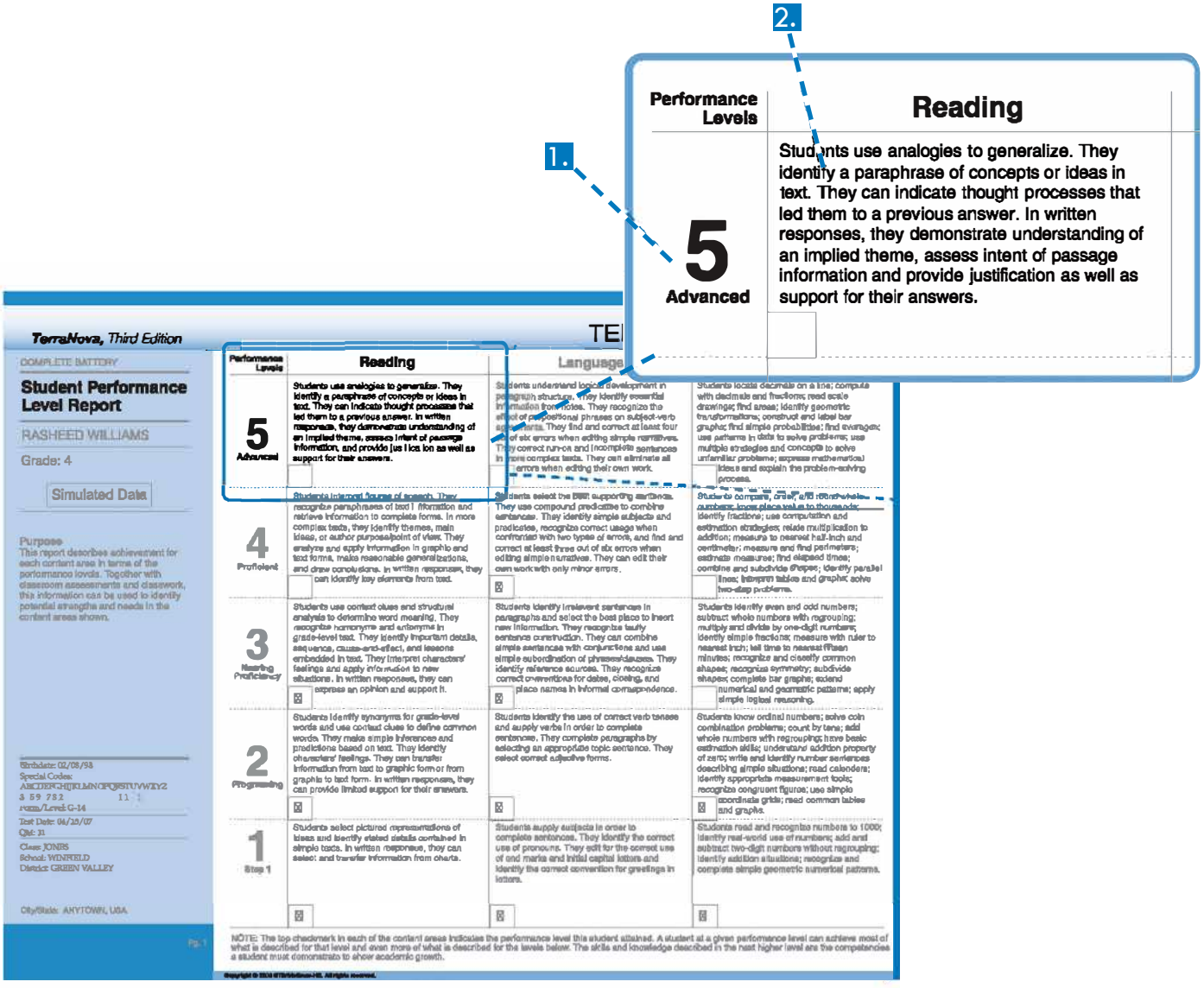
Student Performance Level Reports provide performance level information for individual students. In the sample shown below for a Grade 4 student, performance level scores are shown for three content area tests. Performance Level Reports are also available for classes and other groups or as summary reports.

1. Performance Level Scores

The sample student is Nearing Proficiency in Reading, and is Proficient in Language, which is to be expected, since students should reach the Proficient level by the end of Grade 5. However, he is only at the Progressing Level in Mathematics. This Grade 4 student still has another school year to attain proficiency (Level 4 or Level 5) in each content area. This student will have to work harder in Mathematics than in Reading to reach proficiency.

2. Performance Level Scores

To aid understanding of student performance, DRC provides detailed content-based descriptions of what a student can do at each performance level for each content area. Teachers, students, and parents can use this information to determine what skills and abilities need to be acquired to enable the student to progress to higher proficiency levels.



Glossary of Assessment Terms

Achievement Test
An assessment that measures a student's current acquired knowledge and skills in one or more of the content areas common to most school curricula (for example, reading, language, mathematics, science, and social studies).

Alternative Assessment
An assessment that differs from traditional achievement tests-for example, one that requires a student to generate or produce responses or products rather than answer only selected-response items. This type of assessment may include constructed-response activities, essays, portfolios, interviews, teacher observations, work samples, or group projects. See *Authentic Assessment, Multiple Measures, and Performance Assessment*.

Analytic Scoring
A scoring procedure in which a student's work is evaluated for selected characteristics, with each characteristic receiving a separate score.

Aptitude Test
An assessment designed to predict a student's expected or potential acquisition of knowledge or skills.

Authentic Assessment
An assessment that measures a student's performance in tasks and situations that occur in real life. This type of assessment is closely aligned with and models what students do in the classroom.

Bias
When a test systematically measures differently for various ethnic, cultural, regional, or gender groups, the test is considered biased.

Ceiling
The upper limit of performance that can be measured effectively by a test is called the ceiling. Individuals are said to have reached the ceiling of a test when they perform at the top of the range in which the test can make reliable discriminations. If an individual or group scores at the ceiling of a test, the test is too easy for them, and the next higher level of the test, if available, should be administered.

Constructed-Response Item
An assessment unit with directions and a question or a problem that elicits a written, pictorial, or graphic response is called a constructed-response item. Sometimes it is called an "open-ended" item.

Content Standard
A statement or description of the knowledge and skill in a content area (for example, language, mathematics, science, or social studies) that should be taught in school and assessed in a test is a content standard. National content standards are published by groups such as the National Council of Teachers of Mathematics (NCTM), the International Reading Association (IRA), and the National Council of Teachers of English (NCTE). Local content standards are produced by states and/or school districts.

Criterion-Referenced Test (CRT)
An assessment that measures a student's performance according to specified standards or criteria rather than in comparison to the performances of other test takers is called a criterion-referenced test.

Curriculum-Referenced Test
An assessment that measures what a student knows or can do in relation to specific, commonly taught curriculum objectives is called a curriculum-referenced test.

Derived Score
Any score that is based on (or derived from) a Scale Score is considered a derived score. Examples are National Percentile scores, Normal Curve Equivalent scores, and Grade Equivalent scores.

Distractor
An incorrect answer choice in a selected response or matching test item is called a distractor. Sometimes it is called a "foil."

Equated Score
A score from one test that is equivalent to a score from another test is called an equated score. Equated scores are usually obtained by administering the two tests to a representative sample of students. Scores from one test are then aligned with scores on the other test using equating analyses.

Floor
The opposite of ceiling, the floor is the lower limit of performance that can be measured effectively by a test. Individuals are said to be at the floor of a test when they perform at the bottom of the range in which the test can make reliable discriminations. If an individual or group scores at the floor of a test, the test is too difficult for them, and the next lower level of the test, if available, should be administered.

Frequency Distribution
An ordered tabulation of individual scores (or groups of scores) that shows the number of persons who obtained each score or placed within each group of scores.

Functional Range
The functional range of a test is the range of grades for which the test can be administered in order to obtain accurate norm-referenced data. For most tests, this range is two grades above or below the grade for which the test was intended.

Grade Equivalent (GE)
A score on a scale developed to indicate the school grade (usually measured in tenths of a year) that corresponds to an average test score. A grade equivalent of 6.4 is interpreted as a score that is average for a group that has completed the fourth month of Grade 6. Grade equivalents do not compose a scale of equal intervals and are not usable in drawing profiles.

Holistic Scoring
A scoring procedure that yields a single score based on overall student performance rather than an accumulation of points is considered holistic scoring. Rubrics are used in holistic scoring.

Item
One of the assessment units, an item is a problem or a question that is included on a test.

Item Bias
When an item on a test systematically measures differently for various ethnic, cultural, regional, or gender groups, it is considered biased.

Item Response Theory (IRT) Pattern Scoring
DRC's default scoring methodology, IRT pattern scoring takes into account the difficulty of the items answered correctly, the ability of the items to discriminate between students who know the content and those who don't, and the student's guessing behavior in determining a Scale Score for the student. IRT pattern scoring provides more accurate results for individual students.

Mean
A measure of central tendency a mean is an average calculated by adding a set of scores and dividing by the number of scores in the entire set.

Median
A measure of central tendency, a median is the middle score in a set of ranked scores.

Mode
A measure of central tendency, a mode is the most frequently obtained score in a test.

Multiple Choice
See Selected-Response Item.

Multiple Measures
Assessments that measure student performance in a variety of ways is said to have multiple measures. These may include standardized tests, teacher observations, classroom performance assessments, and portfolios.

Norm-Referenced Test (NRT)
A norm-referenced test is a type of standardized assessment. It is an assessment in which all students perform under the same conditions, and compares a student or group of students with a specified reference group, usually others of the same grade or age.

Number-Correct or "Raw" Score
The Number of Correct Responses (NCR) is the number of items answered correctly by a student on any given test section.

Open-Ended Item
See Constructed-Response Item.

Percentile Rank
The percentage of scores for a particular group that fall below a given student's score determine the student's percentile rank. For example, a score with a percentile rank of 90 indicates that 90 percent of the scores for that particular group are below that student's score.

Performance Assessment
An assessment activity that requires students to construct a response, create a product, or perform a demonstration is called a performance assessment. Usually there are multiple ways to approach a performance assessment, and more than one correct answer.

Performance Standard
A level of performance on a test established by education experts as a goal of student attainment is considered a performance standard.

Portfolio Assessment
An assessment based on a collection of evidence highlighting a student's performance, often over a period of time. A portfolio assessment can contain a wide variety of information, including norm-referenced test scores, awards, drawings, audio or video tapes, and writings. Sometimes these are best examples; sometimes they are a representative sample, exhibiting a record of changes in performance.

Primary Trait Scoring
A scoring procedure in which a student's work is evaluated for one or more specific traits or dimensions; other traits or dimensions are not scored. For example, a student's writing may be scored on organization, but not on grammar or spelling.

Prompt
An assessment topic, situation, or statement to which students are expected to respond is called a prompt. *See also Stimulus.*

Reliability
The degree to which an assessment yields consistent results over repeated uses.

Rubric
A scoring tool, or set of criteria, used to evaluate a student's test or performance.

Scale
An organized set of measurements, all of which measure one property or characteristic. Different types of test-score scales use different units (for example, Number Correct, Percentiles, and IRT Scale Scores).

Selected-Response Item
A question or incomplete statement that is followed by answer choices, one of which is the correct or best answer. Also referred to as a "multiple-choice" item.

Standard Deviation
Defines a range of scores around the mean within which 68 percent of the cases in a normal distribution of test scores are found. About 95 percent of the cases are found within two standard deviations of the mean.

Standard Error of Measurement (SEM)
Defines a range within which a student's "true score" would fall had that student taken the test numerous times. It is important to understand that all tests have inherent measurement error because they are a sample of student performance at one particular time.

Standardized Test
An assessment with directions, time limits, materials, and scoring procedures designed to remain constant each time the test is given, to ensure comparability of scores. Many standardized tests have norms. All norm-referenced tests are standardized.

Stem
The part of an item that asks a question, provides directions, or presents a statement to be completed.

Stimulus
A passage or graphic display about which questions are asked is called a stimulus.

Test
A device or procedure designed to elicit responses that permit an inference about what the test taker knows or can do.

Test Battery
A set of several tests designed to be given as a unit.

Test Objective
A targeted goal that can be measured by an assessment is called a test objective.

Validity
The degree to which a test measures what it is intended to measure.

Additional Publications That Provide Useful Information

Norms Books

Provided by testing period (fall, winter, spring), norms books provide statistical reviewers with norms tables for converting number-correct scores to Scale Scores and for converting Scale Scores to derived scores.

Teacher's Guide to *TerraNova Third Edition*

This comprehensive guide will help teachers understand the *TerraNova* assessment series, how it was developed, how to plan for and administer tests, the rationale and goals of each content area, and how reports can be interpreted and used to facilitate teaching and learning. Included are detailed explanations of items and how they are applied in each content area at each test level; samples of score reports and how to read them; suggestions for communicating results to students and parents or guardians; and tables that classify items by level, objective, and test.

TerraNova, Third Edition Technical Report

This research report overviews test configurations and related testing materials, discuss content rationale and development procedures, and explain the steps DRC took to ensure that the *TerraNova*, *Third Edition* meets all technical requirements. Topics discussed include assessment design and chronology, test construction, bias reduction, scaling and equating, national standardization, reliability, and validity.

TerraNova Test Coordinator's Handbook

This complete manual explains the full cycle used to administer tests and return them for scoring. It includes information test coordinators need to efficiently and effectively instruct teachers on test administration procedures: unpack, sort, and distribute materials; administer tests; and prepare materials for scoring.

Test Directions for Teachers

Each manual (published by test type and level) offers instructions that guide teachers through the process of preparing for, administering, and completing tests and other materials assigned to each classroom.

Concordance between the National Percentile and the Mean Normal Curve Equivalent

National Percentile	Mean NCE	National Percentile	Mean NCE	National Percentile	Mean NCE
99	99	66	59	33	41
98	93	65	58	32	40
97	90	64	58	31	40
96	87	63	57	30	39
95	85	62	56	29	38
94	83	61	56	28	38
93	81	60	55	27	37
92	80	59	55	26	36
91	78	58	54	25	36
90	77	57	54	24	35
89	76	56	53	23	34
88	75	55	53	22	34
87	74	54	52	21	33
86	73	53	52	20	32
85	72	52	51	19	32
84	71	51	51	18	31
83	70	50	50	17	30
82	69	49	49	16	29
81	68	48	49	15	28
80	68	47	48	14	27
79	67	46	48	13	26
78	66	45	47	12	25
77	66	44	47	11	24
76	65	43	46	10	23
75	64	42	46	9	22
74	64	41	45	8	20
73	63	40	45	7	19
72	62	39	44	6	17
71	62	38	44	5	15
70	61	37	43	4	13
69	60	36	42	3	10
68	60	35	42	2	7
67	59	34	41	1	1

DATA RECOGNITION
DRC
CORPORATION

Data Recognition Corporation
13490 Bass Lake Road,
Maple Grove, MN 55311
U.S.A.

www.datarecognitioncorp.com

TN3BTN_120717