NWEA
Northwest Evaluation Association

NWEA FAQ

RIT Scale Norms

Introduction

Northwest Evaluation Association™ (NWEA™) conducts norming studies every several years to provide the best and most up-to-date information we can about student achievement and growth to better support educational decision-making. It is an important part of our commitment to our partners. The most recent NWEA norms were released in July 2015.

Just as we expect student performance to change with time, norms of student performance will change correspondingly. For 2015, several factors may have influenced how the US school-age population performed on Measures of Academic Progress® (MAP®) assessments of language usage, mathematics, and reading over the 2011–12, 2012–13, and 2013–14 school years. Prominent among these possible factors are the introduction of Common Core instruction and Common Core versions of MAP.

But as with all empirical studies, changes in scope and data necessitated changes in statistical design and analyses. The methodological improvements introduced in the 2015 norms have resulted in a set of norms that are more representative of the US school-age population than previous NWEA norms. Methodological changes involved differences in the way post-stratification weights were developed and the way growth was modeled. Details of these changes are addressed more thoroughly in the 2015 Norms Study. As a result of the improved methodology, partners can expect some differences from previous norms, as explained below.

Contents

NWEA

Key Differences from the 2011 Norms

1. What grades and subjects are included in the 2015 norms—are there any changes from prior years?

There are no longer norms for "Science Concepts & Processes," and general science norms for grades 9 & 10 are available only in the 2015 Norms Study.

Please see the chart below for the 2015 norms available by grade and subject.

Table 2.4: 2015 RIT Norms Content Area Coverage

	Grade											
Content Area	K	1	2	3	4	5	6	7	8	9	10	11
Reading	△	△	△	△	△	△	△	△	△	△	△	□
Language Usage			△	△	△	△	△	△	△	△	△	□
Mathematics	△	△	△	△	△	△	△	△	△	△	△	□
General Science				△	△	△	△	△	△	□	□	

△ Status and Growth Norms □ Status Norms only

2. Why don't some grades have general science norms?

Norms for 2nd grade general science are not included. Grade 2 students are tested in science much less frequently then they are for reading, mathematics, and language usage. Given these low testing and participation levels, the grade 2 sample would not be representative of all grade 2 students.

General science status norms for grades 9 and 10 are available for screening and placement decisions pertaining to student attainment in general science skills and concepts. These norms are available in the 2015 Norms Study, and will also be available in Web-Based MAP reports by Fall 2016.

The general science test results should not be used to evaluate performance in topically differentiated high school science courses such as biology, chemistry, or physics, where science content is more specialized.

3. Which grades have status and growth norms?

Status norms (percentiles) are available for the following subjects and grades:

Reading: grades K – 11
Mathematics: grades K – 11
Language usage: grade 2 – 11
General science: grade 3 – 10

Growth norms are available for the following subjects and grades:

Reading: grade K – 10
Mathematics: grade K – 10
Language usage: grade 2 – 10
General science: grade 3 – 8

All subjects and grades for which growth norms were available in 2011 are provided in 2015, with the exception of "Science Concepts and Processes," which have been incorporated into the general science assessment.

Growth norms are now offered for six term pairings:

Fall-to-Winter
Fall-to-Spring
Fall-to-Fall
Winter-to-Spring
Winter-to-Winter
Spring-to-Spring

The inclusion of the winter-to-winter term pair is new to the 2015 norms, and is provided for K – 9 in mathematics and reading, 2 – 9 in language usage, and 2 – 8 in science, where the grade number refers to the student's enrolled grade during the first term of the pairing (e.g., the first winter).

4. What changes will I see in the status percentiles?

The average RIT score associated with a grade tends to be lower in 2015 than in 2011. In addition, the 2015 status norms show greater variance around their means than the 2011 status norms. The combined effect of these differences is that the percentile distributions, particularly in grades 6 and above in mathematics, have shifted up. Thus, the same RIT score for the same grade level, for the same season, and in the same content area is associated with a higher status percentile rank in the 2015 norms than in the 2011 norms.

5. What may have caused the norms to change from 2011?

We observed slight differences from the 2011 norms, some of which reflect true change in the performance of the students. That is to be expected. Conditions in education change over time, and one would expect student achievement and growth to be affected by these changes. Plausible sources for these differences include:

Changes in student demographics

- Student demographics have evolved between 2011 and 2015, with more minority, high-poverty, and English language learner (ELL) students entering schools.

Changes in the educational landscape, such as

- the implementation of the Common Core standards and the associated assessments in many states. In particular, this implementation was asynchronous; that is, schools adopted the assessments at different times and at different paces
- the aftermath of the last recession, which resulted in the largest layoff of teachers since the Great Depression
- the implementation of Race to the Top, which more closely tied teacher evaluation and school accountability to student performance on tests
- improvements to the norming process
- a larger and more diverse student population, from which a more representative sample could be drawn
- refinements to the model for estimating growth that included changes to better estimate summer loss
- moving from five terms of data to nine terms of data
- other methodological improvements

Because we know very little about how these changes impacted individual schools and school districts, we cannot speculate as to what impact they may have on the norms. The norms do represent, however, our best effort to describe the achievement and growth of a nationally representative sample of students between 2012 and 2014.

See Section 1.4 of the 2015 Norms Study for more information about why norms change.

6. NWEA doesn't report RIT scores below 100, so why do the kindergarten norms show first and second percentile math scores below 100?

This is a function of how the model attempts to project the performance of students who are just starting school and have not yet had the opportunity to master the content related to kindergarten. Our reports currently do not report scores below 100, and educators should assume that scores near that level exhibit performance at or near the first percentile.

Typical Growth and Growth Projections

7. How do the 2015 growth norms differ from what we saw in the 2011 growth norms?

The largest noticeable difference between the 2011 and 2015 growth norms is that the newer norms are less "flat" across the achievement distribution. This means that there are larger differences in median (i.e., 50th percentile) growth between students with high versus low initial achievement than we saw in 2011. Put another way, the 2015 norms show higher median growth for students of low achievement than high achievement, whereas under the 2011 norms, median growth differed much less within grades.

8. The status norms at the high school grades are relatively flat across the grades. How can we use the status and growth norms to help track performance and improvement of high school students?

Many high school students show relatively little change in observed MAP reading and mathematics scores over the course of a typical school year. This is particularly true for students with high MAP scores at the beginning of the school year, since, presumably, most such students do not receive instruction in general reading and general mathematics, but take more specialized coursework.

Even so, it is still perfectly appropriate to measure the growth of high school students with the MAP assessment, so long as the focus lies on the change in mean score for the group of students, rather than on individual student gains. This is because the measurement error associated with group means is far smaller than that for individual students, meaning that very small changes in group means over time can be observed with far greater precision than is possible for individual students.

Furthermore, the information from observed MAP scores is still valuable for high school students because it can be used for predictions about likely performance on external measures of achievement such as state summative tests or measures of college readiness.

Changing Percentiles

9. Since the changes in the norms, several students who did not qualify for gifted placement last spring under the old norms now have percentile rankings associated with that same score that would qualify them for gifted placement. What do we do?

Whenever we release new norms, it is likely that the overall distribution of RIT scores will change, meaning that percentile ranks for some students will change as well. Changes in the norms reflect a change in the standard, not a change in student performance. If decisions are made on the basis of points on the score scale (e.g., RIT) instead of percentiles, changes in percentiles that almost invariably accompany changes in the norms will be far easier to explain to stakeholders. We would recommend treating all students who achieve a qualifying score, whether that score was gained under the old norms or the new norms, as crossing the qualifying threshold. And of course we'd recommend that the district consider multiple criteria in addition to MAP scores to determine any program eligibility.

Study Data & Methodology

10. What were the nine terms that were used in the 2015 norms study?

The terms used for the 2015 norms were the fall, winter, and spring terms of the 2011–12, 2012–13, and 2013–14 academic years. 2012–13 was the "focal" year—the year used to report status norms. Data from all three years were used to model change (growth).

11. MAP and MAP for Primary Grades (MPG) are both often taken at second grade. Does that factor into how the norms are reported at that grade?

The norms are intended to describe the achievement and growth of a nationally representative student population in the subjects tested. Because they are intended to show achievement and growth of second graders in mathematics, and because NWEA assessments are aligned to the RIT scale, we do not differentiate MPG and MAP in the norms. If we did so, the second grade norms could not reflect the achievement of a nationally representative sample in the subject.

One clear difference between MPG and MAP is the wide availability of audio help. This adaptation is intended to assure that measurement of a student's achievement in a subject is not confounded by their inability to read questions. For example, we would not want the reading level of a first grader to interfere with their ability to demonstrate what they can do in mathematics. The presence of audio help, when used in accordance with our administration guidelines, improves our ability to accurately measure mathematics for all younger students.

12. What testing windows for fall, winter, and spring were used in the norms? Were the weeks used the same as the 2011 Norms Study?

The 2015 Norms Study shows status norms tables for fall, winter, and spring testing by subject and grade. These tables assume that a fixed number of weeks of instruction have occurred between each testing season (i.e., four weeks for fall, 20 weeks for winter, and 32 weeks for spring). However our reporting systems adjusts these instructional weeks assumptions based on the testing schedules specified for each school, so the printed tables in the norms document may differ from what are shown on school reports, depending on the testing schedules of schools.

13. How was the School Challenge Index (SCI) used in the 2011 and 2015 studies?

The SCI used in both studies was only used in the post-stratification process. The weights referred to were only applied to the schools selected for inclusion in the norm studies. Their purpose was to ensure that the randomly selected sample used was representative of the US school-age population with respect to the collection of school-level variables that can be interpreted as indicators of school challenge. SCI values are not applied to scores reported from ordinary operational tests. The difference between SCI 1.0 (form the 2011 study) and SCI 2.0 (from the 2015 study) is that the version 1.0 was referencing schools in the same state, whereas the version 2.0 references schools across states. The 2015 Norms Study provides more details about the SCI and how the norming sample was created.

14. What is meant by "reduces seasonal bias" in reference to changes to the growth model?

The new method of curve-fitting more faithfully retains the "drop" in scores sometimes seen over summer. The result is that fall scores will be less likely to be over-predicted and spring scores will be less likely to be under-predicted. Therefore, Fall–Spring growth is greater in the 2015 growth norms than it was in the 2011 growth norms (more so in the lower grade levels than upper).

School Norms

15. Can I use the school norms tables to calculate the district status norms?

That would not be advisable, as there are no district-level percentile ranks. Using the displayed percentiles to make inferences about a higher level of aggregation will not lead to a meaningful or a valid inference. Just as making inferences about schools from student-level norms would lead to invalid inferences about schools, making inferences about districts from school-level norms would also lead to invalid inferences about districts.

It would be informative to look at a picture/graph of all the schools in terms of their individual performance, either as School CGI (Conditional Growth Index) or School Conditional Growth Percentile (CGP). This would not reflect the district per se, but would give a picture of how the district, as a collection of schools, performed.