K-12 Research Idaho

December 23, 2016

The IRI and the Idaho Reading Initiative

Bert D. Stoneberg

The State Department of Education is responsible for collecting and reporting results from the Idaho Reading Indicator (IRI). Staff in local schools record IRI "correct scores" and skill level (i.e., achievement level) for each student in kindergarten through the third grade in the fall and in the spring. Figure 1 shows the fall IRI scores recorded for third graders.

3rd Grade **Idaho Reading Indicator**
Fall **Student Record Card Grades K-3**

Subtest	Total Correct	Skill Level
Reading CBM (RCBM)		
Reading CBM Total		
RCBM Score Determines Skill Level		

RCBM	3	77 & Above
	2	49 - 76
	1	0 - 48

Figure 1. The IRI correct score and skill level is recorded on the Student Record Card for each student fall and spring.

The Department has elected to report only the number of students and the percentage of students at each IRI skill level. The three IRI skill levels are Benchmark (3), Strategic (2), and Intensive (1). Figure 2 shows what the Department reported for the fall 2015 IRI results for third graders.

Idaho Reading Indicator
3rd Grade Fall 2015

Score	*Number Of Students*	*Percent Of Total*
Intensive	3,382	14.76%
Strategic	4,888	21.34%
Benchmark	14,640	63.90%
	22,910	

Source: Idaho State Department of Education

Figure 2. The number and percent of third graders at each skill level (i.e., achievement level) reported for the fall 2015 IRI assessment.

There has been at least one news report recently about a "wide reading gap" existing in Idaho schools based on IRI achievement level counts and percentages that the Department has released (Richert, 2016). IRI achievement level results do not support this claim for two reasons. First, achievement levels distort the meaning of large "correct score" differences within a single level and small differences across neighbor achievement levels. For example, the student with the highest score in the state is treated exactly the same level of achievement as the student who barely met the cut-score for the Benchmark level. Moreover, the student who barely met the Benchmark cut-score has demonstrated grade level skill, but the student who scores just one point below that cut-score has not.

Second, whoever first demonstrates that achievement level results are "reasonable, valid, and informative to the public" surely will be famous forever. According to the National Academy of Sciences, methods for setting cut-scores by groups composed of members qualified and competent in psychometrics, curriculum, instruction, child development and statistics are fundamentally flawed (i.e., cannot be fixed) and are not scientific (i.e., cannot be replicated) (Pellegrino, et al, 1999).

What could the state learn from IRI with appropriate data collection, analysis and reporting?

Skill Levels. Figure 3 displays the statewide fall IRI kindergarten skill-level-counts for the 2014-15, 2015-16, and 2016-17 school years. Take some time to write down a couple of statements describing the reading skills of students entering Idaho schools over these three years. Invite someone else to do the same, and then compare your observations. Find much agreement?

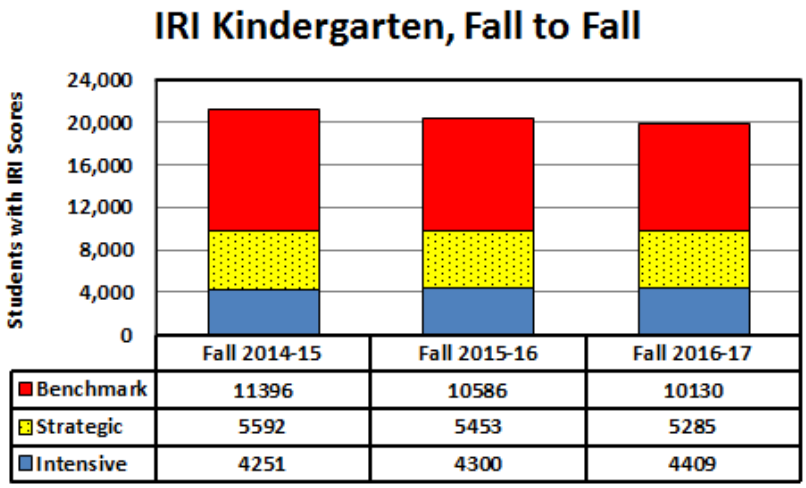


Figure 3. Student count by skill level on the kindergarten fall IRI for the 2014-15, 2015-16, and 2016-17 school years.

Learning Point Associates (2009) has suggested three measures of student achievement -- once in widespread use but ignored today -- provide program evaluation and accountability data more useful than the percentage of students in this or that achievement level. The three measures include averages (means), percentiles and effect sizes. Percentiles and effect sizes are calculated using the number of students, the mean score, and the standard deviation of the students tested.

The state does not collect the IRI student scores needed to calculate the mean score and standard deviation for any group be it the state, a district or a school. There are, however, statistical procedures to estimate Idaho's mean and standard deviation of from the data grouped by the three IRI skill levels (Fonolahi, 2014). These are estimates because there is no way to recover all of the information lost when more than 20,000 scores ranging over 75 or 100 points are reduced to three arbitrary points, namely the mid-score of the three skill levels. The number of students, the estimated means (averages) and the estimated standard deviations for the fall kindergarten IRI were 21,239, 10.092, and 0.530 in 2014; 20,339, 9.985, and 0.559 in 2015; and 19,824, 9.754, and 0.579 in 2016. These estimates are interesting because they describe the reading skills of students as they enter Idaho's schools.

Average. The difference between two means can have *statistical significance* and *practical significance*. The *student t-test* is the probably the most widely used procedure for describing the level of statistical significance between two means.

Effect size. Effect size is a commonly used indicator of practical significance that describes the *magnitude* of the difference between two means. It is calculated from the means and standard deviations. Figure 4 shows the estimated effect sizes on the fall kindergarten IRI for 2014-15 to 2015-16, and for 2015-16 to 2016-2017. The invitation from Exhibit 3 is renewed. Take some time to write down a couple of statements describing the reading skills of students entering Idaho schools over these three years. Invite someone else to do the same, and then compare your observations. Find much agreement?

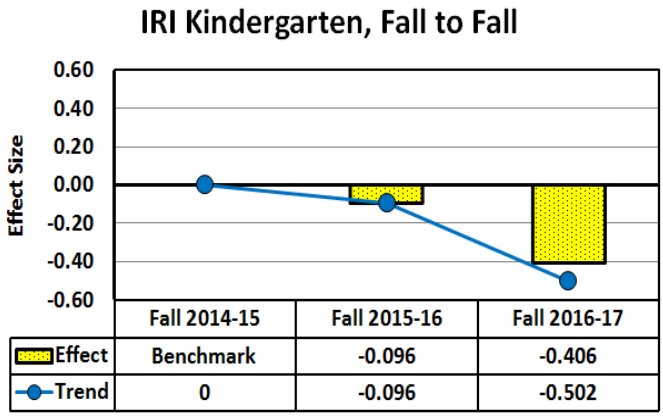
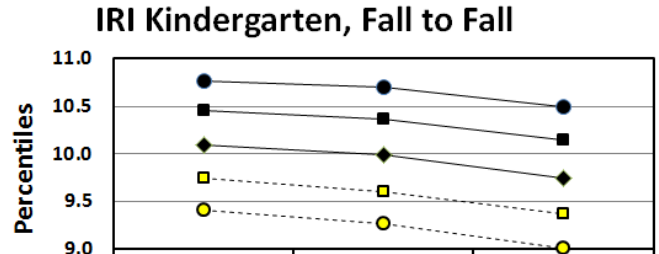


Figure 4. Effect size (magnitude) of the difference of fall kindergarten IRI estimated mean scores from 2014-15 to 2015-16 and from 2015-16 to 2016-2017, by year and by trend.

Percentiles. The mean can hide a lot of change across the student distribution. Percentiles, on the other hand, enable us to identify changes at five points across the distribution: high student (90th percentile), high average student (75th percentile), average student (50th), low average student (25th), and low student (10th). From assessment to assessment, the high student is compared with the high student, the average student with the average student, etc.

Percentiles are calculated from the mean and standard deviation. They may be displayed as either as a line graph (when detail is desired) or a box plot (when an overall impression is desired). The estimated percentiles from the fall kindergarten IRI from 2014-15, 2015-16 and 2016-2017 are displayed in Exhibit 5 (line graph) and Exhibit 6 (box plot).



	Fall 2014-15	Fall 2015-16	Fall 2016-17
90th	10.77	10.70	10.50
75th	10.45	10.36	10.14
50th	10.09	9.99	9.75
25th	9.74	9.61	9.37
10th	9.41	9.27	9.01

Figure 5. Line graph and estimated scores of the 90th, 75th, 50th, 25th and 10th percentiles on the fall kindergarten IRI from 2014-15, 2015-16, and 2016-17.

In the box plot, the middle half of the students are "in the box." Ten percent of the students are above the high score, and ten percent are below the low score.

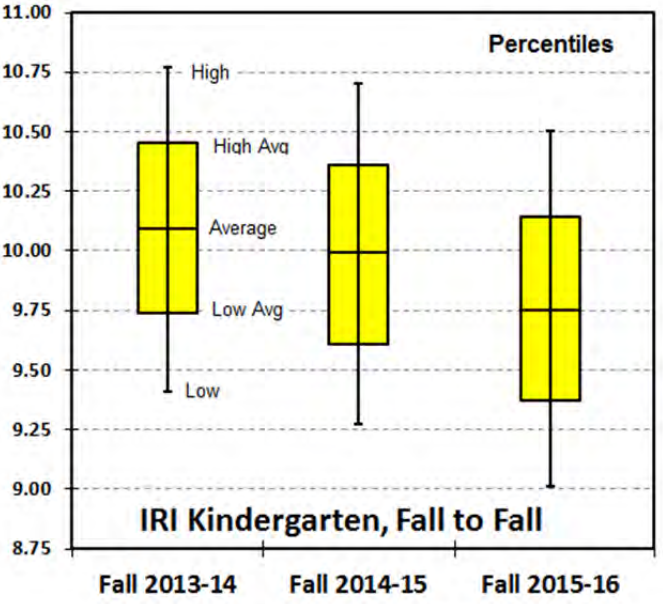


Figure 6. Box plot of the 90th (high), 75th (high average), 50th (average), 25th (low average) and 10th (low) percentiles as estimated for the fall kindergarten IRI from 2014-15, 2015-16, and 2016-17.

Some might think that IRI data displayed figures 4 and 6 might provide evidence to support statewide pre-K programs in Idaho. It must be emphasized, however, that the estimated effect sizes and percentiles displayed above may have been smaller or larger had the means and standard deviations from which they were calculated had actually been based on all Idaho student IRI scores rather than the estimate illustrated in this whitepaper.

What IRI data does the SDE need to collect to evaluate the Idaho Reading Initiative?

At a minimum, the SDE should be able to report the number of students with valid test scores, the mean for these valid scores, and the standard deviation of these valid scores for each IRI assessment. The SDE should use these three IRI statistics to examine the statistical significance (t-test) and the practical significance (effect size) of each assessment. The SDE should also examine the differences in IRI percentiles from assessment to assessment.

The minimum statistics in the list above should be collected for the state and for the student groups in which the state has interest with regard to their reading skills. These groups of interest may include gender (male, female), ethnicity (White, Hispanic, American Indian, etc.), poverty (eligibility for school lunch program), etc.

In fact, the SDE should collect student level data and report a basic set of statistics for each assessment that the State Board of Education requires statewide student participation:

Number of Students
Mean (Average)
Standard Deviation
Effect Size (difference between two means)
Percentiles: 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th}

# # #

References

Fonolahi, A. (2014). *Calculate standard deviation using group data*. Presentation available online at https://www.youtube.com/watch?v=abYUPzRMzcQ

Learning Point Associates. (2009, August). *Connecting Research to Practice: Knowing Who Is Proficient Isn't Always Sufficient* . Retrieved July 14, 2016, from http://files.eric.ed.gov/fulltext/ED509965.pdf

Pellegrino, J., Baxter, G.P., Bradburn, N.M., et al. (1999). Grading the nation's report card: evaluating NAEP and transforming the assessment of educational progress. *National Academy Press*. Available online at https://www.nap.edu/read/6296/chapter/8

Richert, K. (2016, Dec. 14). Idaho schools try to bridge a wide reading gap. *Idaho Education News*. Available online at https://www.idahoednews.org/news/idaho-schools-try-bridge-wide-reading-gap/

Citation: Stoneberg, B.D. (2016, December 23). *The IRI and the Idaho reading initiative*. Available online at http://k12researchidaho.com/pdf/iri-and-readinginitiative.pdf