SYLVAN
RESEARCH
I N S T I T U T E

The Case for the Use of the STAR
in Sylvan Centers in Canada

September, 2013

The Case for the Use of the STAR in Sylvan Centers in Canada

Prepared by
Richard C. Gershon, Ph.D.

About the Author

Dr. Richard Gershon is Vice Chair for Research in the Department of Medical Social Sciences at the Feinberg School of Medicine at Northwestern University. Dr. Gerhson is a leading expert in the application of Item Response Theory (IRT) in individualized and large-scale assessments. He has developed item banks and Computerized Adaptive Testing (CAT) for education, clinical, and health - including cognitive, emotional, and motor applications.

He currently is the principal investigator overseeing field operations for the National Children's Study (NCS) Southern Regional Operating Center. He also chairs the overall NCS Health Measurement Network to develop a measurement strategy for 100,000 children, and their parents, from birth to age 21. He is also the principal investigator for the National Institute's of Health (NIH) Toolbox for the Assessment of Neurological and Behavioral Function and for the NIH Common Fund Patient Reporting Outcomes Measurement Information System (PROMIS) Technical Center.

STAR in Canada

EXECUTIVE SUMMARY

Sylvan Learning selected the STAR Math and Reading Enterprise Assessments to aid in initial placement and later evaluation of progress for their students. The STAR tests are already used in some Canadian schools, even though Canadian-specific norms were never produced. In an effort to examine the potential impact of using primarily U.S.-derived norms for Canadian students, we conducted three different analyses. First we reviewed three major studies that examined differences in student progress between the U.S. and Canada: the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). In addition, in 2010, we used data provided by Renaissance Learning to compare the U.S. norms with the data collected from all students using the STAR series in Canada. We then compared the STAR results for the Canadian students who had taken the test with the results of U.S. students. Finally we completed an alignment of Canadian provincial standards to U.S. Common Core Standards and the Sylvan Curriculum.

Using the international results, we found that overall between-country differences were fewer than the observed differences among states or provinces within each country. When we conducted an analysis of 65,523 STAR tests administered in Canada, we found that generally speaking, mean student performance closely matched the U.S.-normed-based grade equivalents for Grades 1–6, with minor divergence for Grade 7 and above.

BACKGROUND

In 2009, Sylvan Learning issued a Request for Proposals to major test publishers to provide reading and math assessments that were suitable for both norm-referenced testing (NRT) and criterion-referenced testing (CRT) purposes and that would be suitable for use by Sylvan Learning Centers in both the United States and Canada. In 2010, Sylvan conducted extensive analyses of the proposal made by Renaissance Learning (RNL) to provide their STAR Math and STAR Reading tests to fulfill this NRT need and found that the Renaissance proposal was the strongest. This analysis was followed by an extensive preliminary study in which STAR Math and Reading were used in a set of Sylvan Centers, in addition to the CAT/5 and 2, in order to understand the differences as well as the potential that the STAR tests presented. This paper provides an overview of the specific analyses that were conducted to determine the appropriateness of the STAR for use in Canada.

Sylvan Centers in Canada have reported that some of their students take the STAR tests in their home schools. Indeed, in 2010, STAR Reading was administered to over 29,000 students across Canada. STAR Math was administered to significantly fewer students. RNL score reports in Canada present an absolute ability score as well as norm-referenced information (such as grade equivalents) based upon the same norming sample used to produce reports for U.S. schools (a sample that did not include Canadian schools).

It was logical to first determine whether or not there are differences in ability between Canadian and U.S. students by looking at other international tests that compare student achievement. A review of the literature found that there are three international testing programs that are typically utilized to compare student progress among different countries: the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS).

PISA

"The Program for International Student Assessment (PISA) is an international assessment that measures 15-year-old students' reading, mathematics, and science literacy. PISA also includes measures of general or cross-curricular competencies, such as problem solving. PISA emphasizes functional skills that students have acquired as they near the end of compulsory schooling.

PISA is coordinated by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of industrialized countries and is conducted in the United States by NCES. PISA was first administered in 2000 and is conducted every three years. The most recent assessment was in 2012." [1]

In 2000, the focus was on reading literacy. U.S. students scored an average of 504, while the Canadian average score was 534. The worldwide average is fixed at 500. The reading test was repeated in 2009, when the average worldwide score dropped to 493, the U.S. average dropped to 500, and the Canadian average dropped to 524. None of the drops were deemed to be statistically significant.

In 2003, the mathematics examination was given. U.S. students scored an average of 483, and Canadian students scored an average of 532. The 2009 repetition of the test found that U.S. student scores improved to an average score of 487, while Canadian students retreated to an average score of 527. None of the differences between 2003 and 2009 were found to be statistically significant.

TIMSS

"The Trends in International Mathematics and Science Study (TIMSS) 2011 is the fifth in IEA's series of international assessments of student achievement dedicated to improving teaching and learning in mathematics and science. First conducted in 1995, TIMSS reports every four years on the achievement of fourth and eighth grade students. A number of countries participating in TIMSS 2011 will have trend data across assessments from 1995 to 2011." [2]

Canada had full participation to evaluate grade 4 and grade 8 performance in 1995, and then only for grade 8 students in 2003 and again in 2011. Most provinces have been represented individually across this timeframe. The United States fully participated at all three time points. In 2011, U.S. students (grade 4 = 541, grade 8 = 509) scored above the international mean in math performance. U.S. grade 4 performance improved from an average of 529 in 2007 to 541 in 2011. Scores in Ontario and Alberta showed similar improvements between 2007 and 2011 as well.

[1] "Program for International Student Assessment (PISA)." National Center for Statistics. U.S. Department of Education, Institute of Education Sciences, n.d. Web. 18 Sept. 2013. [http://nces.ed.gov/surveys/pisa/]

[2] "About TIMSS 2011." TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. International Association for the Evaluation of Educational Achievement, n.d. Web. 18 Sept. 2013. [http://timssandpirls.bc.edu/timss2011/index.html]

Table 1. TIMSS Distribution of Mathematics Achievement - Grade 4	Average Scaled Score	
	2007	**2011**
United States	529	541
Massachusetts	547	-
Minnesota	532	-
Ontario	512*	518
Alberta	505	507

*Note: Ontario students were on average one-half grade younger than their U.S. counterparts at the time of testing.

Table 1 illustrates several things that are important to consider. First, while the Canadian provinces lag behind the U.S. overall, the average age at time of testing was less, and therefore one would expect scores to be slightly lower. Another interesting observation is made in comparing the United States' overall average to the two specific states. Students in both Massachusetts and Minnesota outperformed the United States' overall average. This discrepancy helps to underscore the caution that must be associated with citing any national "norm." Students from these states will appear to outperform the U.S. average, and hence even grade equivalents will appear to be high.

Table 2 shows average grade 8 results for TIMSS mathematics achievement. Once again, mean scores for most of the participating Canadian provinces exceeded the U.S. overall average, but so did the mean scores for the two individually participating states.

Table 2. TIMSS Distribution of Mathematics Achievement - Grade 8	Average Scaled Score	
	2007	**2011**
United States	508	509
Massachusetts	547	561
Minnesota	532	545
Ontario	528	512
British Columbia**	509	-
Alberta	-	505

** No Grade 4 data was available for British Columbia.

PIRLS

The Progress in International Reading Literacy Study (PIRLS) is an international comparative study of the reading literacy of young students. PIRLS studies the reading achievement and reading behaviors and attitudes of fourth-grade students in the United States and students in the equivalent of fourth grade in other participating countries.[3] PIRLS was most recently administered in 2011 to grade 4 students in 53 education systems (including countries and subnational education systems, such as Canadian provinces). In particular, scores are calculated for Combined Reading Literacy, Literacy Subscale, and an Informational Subscale. Table 3 shows the average combined literacy scores for students in the U.S. and select Canadian provinces.

Table 3. Progress in International Reading Literacy Study (PIRLS)	Average Combined Reading Literacy Score	
	2006	2011
United States	540	556
Ontario	555	555
British Columbia	558	-
Alberta	560	548
Nova Scotia	542	-
Canada	-	548

The PIRLS results once again highlight the degree of variability within a country, and bring into question the true value of using a single set of national norms or grade equivalents (although indeed all but one province did produce mean scores that exceeded the overall U.S. average in 2006, but not in 2011). The addition of a total score for Canada was added in 2011, which was less than that for the overall U.S. average.

In looking across all of the international studies, it can be argued that U.S. and Canadian performance is quite similar. Grade 4 math scores were similar for both the U.S. and Canadian students—any under-performance in the Canadian provinces was likely due to the slightly lower ages at which the tests were administered in Canada. Average U.S. math performance was much lower than that observed specifically in Massachusetts and Minnesota, which were both higher than that observed in Ontario—again any such differences were possibly due to small variations in age at time of testing. Grade 4 reading results were cited for the total U.S. sample versus five Canadian provinces. Again, the results are quite similar (two provinces scoring similarly or below the U.S. average, while three provinces were higher). Clearly, differences across the border are generally small, and vary quite widely by state and province.

"PIRLS 2011 International Results in Reading." *TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.* International Association for the Evaluation of Educational Achievement, n.d. Web. 18 Sept. 2013. [http://timssandpirls.bc.edu/pirls2011/international-results-pirls.html]

THE STAR ASSESSMENTS FROM RENAISSANCE LEARNING

The STAR series has only been normed on a large U.S. sample. In 2010, using data provided by Renaissance Learning, we compared the U.S. norms with the data collected from all of the students using the STAR series in Canada. Initial analyses compared all 2008–2009 Canadian students to the U.S. norms, and to a reduced sample consisting only of students taking the test in the spring. Finally, in order to best match the sample utilized to produce the 2008 U.S. norms, the Canadian sample was limited to students who tested between April 15 and May 30, 2009 without special administrations (e.g., extended time or with a reader). Furthermore, the Canadian sample used only the first test administered during this time period (but not if it was a retake of a test taken prior to April 15). Two comparisons were made available. The first showing the U.S. "unweighted" sample—unweighted in that it included all of the RNL students who took the reading test for the first time during this time period irrespective of the fact that those students were not representative of the U.S. general population. This can be viewed in contrast to the U.S. "weighted" sample (the result of RNL's effort to weight the results obtained using their "convenience" sample of all U.S. STAR students to better represent the scores that would have been obtained had true U.S.-wide random sampling taken place). Weights were applied to correct for region (Midwest, Northeast, Southeast, Western), race/ethnicity (white, black, Hispanic, Asian/Pacific-Islander, other), and gender.

These analyses were updated for the July 2013 version of this report. Renaissance Learning aggregated the data for all of the students completing the tests in Canada from August 1, 2009 through June 12, 2010—a total of 65,523 test records. Of these records, 1,136 testers used extended time, and 3,894 testers ran out of time and did not complete 1–4 items. A total of 188 Canadian schools are included, but many of them with limited participation.

Generally speaking, mean placement grade level at time of testing matches the U.S.-normed-based grade equivalent for Grades 1–6 (see Table 4). These numbers begin to diverge from grade 7 and up (likely for grades where STAR is primarily used to identify the need for remediation as opposed to testing of all students).

Table 4. Mean Placement versus U.S.-Norms-Derived Grade Equivalent for Canadian Students				
Total Sample				
Grade	N	Mean Placement At Time of Testing	Mean Test Result Grade Equivalent	Difference
1	2073	1.60	1.54	.06
2	6477	2.45	2.35	.10
3	9022	3.42	3.31	.11
4	9368	4.41	4.25	.16
5	9851	5.39	5.17	.22
6	7942	6.36	6.05	.31
7	6664	7.37	6.69	.68
8	6376	8.37	7.64	.73
9	4105	9.31	8.02	1.29
10	1641	10.37	8.33	2.04
11	1130	11.39	9.64	1.75
12	1826	12.36	10.01	2.35

This relationship can also be observed graphically. Each point represents the average score obtained at a given school (see Figure 1).

Figure 1. Grade Placement Compared to Test Performance Reported as a Grade Equivalent
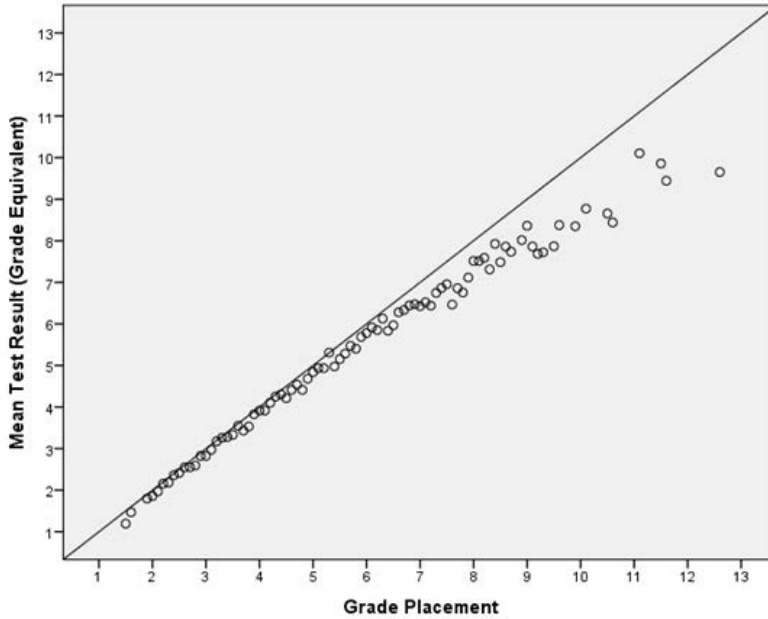
In addition to the fact that there appears to be almost no mean difference in Canadian versus U.S. student performance on the STAR, there is at least one additional reason to rely on the U.S.-derived norms. Several Sylvan operators in Canada have reported that their students already use the STAR tests. Parent and school-level reporting of these tests presumably includes results based on the U.S.-sample derived norms. This single factor would strongly lead to the conclusion that Sylvan should not provide different normative information than is already being provided to students by their schools. Any attempt by Sylvan to produce a grade equivalent or percentile norming system that differs from what schools are distributing to their students, for the same test, would be seriously counterproductive and presumably confusing for parents.

COMMON CORE, SYLVAN CONTENT AND PROVINCIAL STANDARDS

The Common Core State Standards are a set of international benchmarks that have been adopted by a significant number of states. Renaissance Learning has used the data from millions of STAR test results to create empirically validated learning progressions that are mapped to Common Core State Standards. Sylvan mapped its content to these progressions in reading and math, creating its own progressions that are used to systematically organize content in a coherent fashion. Sylvan's Common Core aligned learning progressions are key to the design of SylvanSync. The results of Sylvan's skills assessments are used to place students at a starting point on the appropriate Sylvan progression. Once on the progression, the student's experience with the content and ability to achieve mastery determine which content a student receives when and in what order, creating an adaptive learning experience.

Sylvan Learning has produced alignments of select Canadian provincial standards to Common Core State Standards and related SylvanSync skills using the Academic Benchmarks alignment tool. These documents demonstrate the applicability of Common Core to the provincial standards. The alignments can be found on SLCLlnk>Education> Alignments>Canada. Other information on the Common Core and the SylvanSync content can also be found at SLCLink>Education>Common Core.

CONCLUSIONS

An examination of several international studies found that overall scholastic differences between the United States and Canada were often fewer than observed differences among individual states or provinces within the two countries. We also found that even though the norms used to interpret STAR tests delivered in Canada are actually based on a U.S. sample, the mean student performance closely matched the U.S.-normed-based grade equivalents for grades 1–6, with only minor differences observed for grade 7 and up. In addition, the STAR tests are specifically designed to measure growth toward academic competency as defined by the Common Core State Standards, which are internationally benchmarked standards. As is illustrated in the alignments of Canadian provincial standards to the Common Core State Standards and Sylvan content located on SLClink>Education>Alignments, the STAR tests are an excellent choice for U.S. and Canadian centers.