Athens Journal of Education - Volume 3, Issue 4 – Pages 302-312

An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study

By Jianjun Wang[*]
Xin Ma[†]

The United States invented a matrix sampling technique to impute five plausible scores of student performance in its National Assessment of Educational Progress. That approach was adopted by a Trend in Mathematics and Science Study (TIMSS) for international comparison. In this paper, baseline data from TIMSS 1995 are analyzed at the seventh grade to examine correlation of plausible scores between mathematics and science. Canonical correlation is introduced to address a non-additive nature of correlation coefficient and reduce Type I error in the result aggregation. Besides revealing the impact of interdisciplinary correlation, this investigation reconfirms importance of student performance in each core subject.

Keywords: canonical correlation, math & science achievement, international comparison.

An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study

In 2014, 46 states have adopted Common Core Standards and two additional states were moving toward the same direction (Fensterwald, 2014). As a result, school districts were given the authority to change mathematics and science curricula in K-12 education (Brugger, 2014; Will, 2014). Meanwhile, National Science Foundation (NSF) integrated its funding across science, technology, engineering and mathematics (STEM) education (NSF, 2013). Alphonse (2014) predicted that "STEM skills may be required in as many as 50% of future jobs" (p. 15).

To assess learning outcomes in each core subject, National Assessment of Educational Progress (NAEP) produced report cards in mathematics and science for more than four decades (Jacob & Ludwig, 2009). However, mathematics and science scores were gathered from different students. Thus, no national indicators can be developed to correlate mathematics and science achievements in NAEP (Johnson, 1998).

Instead of reinventing a wheel to fill this void, it was designed in this study to borrow the wheel, i.e., using the Trend in Mathematics and Science Study (TIMSS) to construct a correlation indicator between mathematics and science achievements in international context. Although Program for International Student Assessment (PISA) is another well-known project, similar correlation

[*] Professor, California State University, USA.
[†] Professor, University of Kentucky, USA.

https://doi.org/10.30958/aje.3-4-1 doi=10.30958/aje.3-4-1

analyses cannot be conducted on PISA because the assessment data are collected in one subject each year. Hence, TIMSS is the only large-scale study that concurrently includes both mathematics and science assessments in its data gathering.

To date, no one correlated STEM performance in TIMSS. While researchers attempted to examine 4th grade performance across disciplines (Martin & Mullis, 2013), their report was focused on reading demand in TIMSS testing, and did not include a correlational study of academic performance between mathematics and science.

In the United States and many western countries, TIMSS researchers reported that mathematics and science curricula were "one mile wide and an inch deep" (Jakwerth, 1996; Schmidt et al., 2001). While lengthy tests were needed to cover the broad curricula, a time-consuming test is not feasible for TIMSS assessment. As a compromise, TIMSS designated booklets to split the entire tests in mathematics and science, and each student answered questions in one booklet (Foy, Arora, & Stanco, 2013). Thus, matrix-sampling techniques were introduced to impute five plausible scores of student performance in mathematics and science (Cogan, 1996). The imputed scores were aggregated for dissemination in international report cards (Beaton et al., 1996a; 1996b).

Under an assumption that countries can learn from each other, this study is designed to construct correlation indicators of student performance across nations. Davison, Miller, and Metheny (1995) noted, "Whether the integration of science and mathematics occurs within the disciplines or is infused with the disciplines, integration will provide for a more reality-based learning experience" (p. 229). Vygotsky (1987) also acknowledged links of mathematics and science performance in school settings. Lewis, Alacaci, O'Brian, and Jiang (2002) further maintained that mathematical knowledge can improve the quality of scientific inquiry. Therefore, correlation indicators are needed to assess the impact of interdisciplinary connection on STEM learning in a cross-country context.

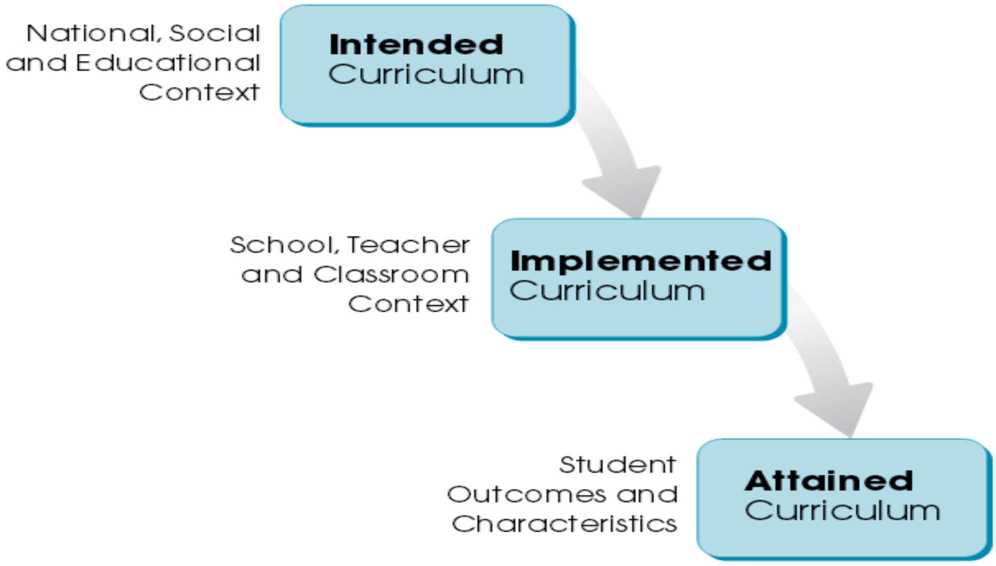
Literature Review

During the Cold War, mathematics and science education was linked to national security (Rotberg, 1990). To assess the learning outcome, the first and second international mathematics studies were launched during 1960s - 1980s by the International Association for the Evaluation of Educational Achievement (IEA) (see Wang, 1998; 2011). Although Rotberg (1991; 1995) questioned credibility of the comparative data, Bradburn, Haertel, Schwille, and Torney-Purta (1991) noted that Rotberg "sees improvement in the switch from the First International Mathematics Study (in which all students in the final year of secondary school were sampled) to the Second International Mathematics Study (in which only those students still taking mathematics were samples)" (p. 775). In the 1990s, additional improvement was made in IEA studies to include the matrix sampling technique in the Third International

Mathematics and Science Study (TIMSS). Subsequently, the TIMSS acronym was redefined as "Trends in Mathematics and Science Study" because of repetitions of the IEA study every four years (Mertens, Anfara, & Roney, 2009).

In support of the ongoing trend study, researchers established a three-tier model to describe *intended*, *implemented*, and *attained* curricula in TIMSS (Mullis & Martin, 2013). While *intended* and *implemented* curricula were based on collaboration of education stakeholders at different levels (see Figure 1), it is the *attained* curriculum that justifies learning outcomes of school accountability in each subject (IEA, 2011).

Figure 1. Three-Tier Curriculum Model



Despite the subject division in mathematics and science, it seems farfetched to assume no interdisciplinary support in the *intended* curriculum. According to Haigh and Rehfeld (1995), "Several organizations such as School Science and Mathematics Association (SSMA), the American Association for the Advancement of Science (AAAS), and the National Research Council have given strong support for the integration of mathematics and science education" (p. 240).

The interdisciplinary link also occurred in the *implemented* curriculum. Johnson (2011) reported that "Being able to teach math better and being able to teach science better are powerful reasons for the math and science teacher collaborate with each other" (p. 1). McBride and Silverman (1991) suggested two reasons for integrating mathematics and science education: (a) "Mathematics can enable students to achieve deeper understanding of science concepts by providing ways to quantify and explain science relationships" and (b) "Science activities illustrating mathematics concepts can provide relevancy and motivation for learning mathematics" (p. 287). Stodolsky and Grossman (1995) concurred that "in sequential subjects, teachers report more coordination with colleagues" (p. 227).

At the *attained* curriculum level, inter-subject correlation can be established from empirical findings. Since some concepts, such as "proportion" in mathematics and "density" in physics, can be identified to connect STEM disciplines (Lester, 2007), the relationship should be ultimately reflected in student performance between mathematics and science. Prior to the first round of TIMSS data collection in 1995, no one gathered large-scale data to support a correlational study between mathematics and science in NAEP or other IEA projects (Wang, 1998). Thus, this study represents an innovative attempt to support construction of correlation indicators for the ongoing TIMSS studies every four years.

In summary, this investigation of mathematics and science education is grounded on a solid theoretical framework of *intended*, *implemented*, and *attained* curriculum. The construction of correlation indicators is designed to fill a literature gap in international studies. Empirical evidence from TIMSS can be employed to reconfirm or disconfirm the impact of inter-subject correlation on science learning outcomes in a cross-country context.

Research Questions

Like NAEP, TIMSS produced mathematics and science reports on a regular basis (Martin & Mullis, 2013). To identify and illustrate a feasible method for computing the indicator of inter-subject correlation, baseline data from TIMSS 1995 are employed in this study to address three questions:

1. What method can be used to correlate plausible scores between subjects without inflating Type I error?
2. Does the approach support result aggregation across plausible scores?
3. Does the correlation index impact student achievement in science at the seventh grade?

These questions have important methodological and policy implications. From the methodological perspective, no one has attempted to solve questions 1 and 2 in the past. Although the American Educational Research Association (AERA) received funding from National Science Foundation and the National Center for Education Statistics to support TIMSS data analyses (see AERA, 2015), researchers need the methodological support to handle non-additive nature of correlation coefficient for grant application. In addition, Hellier (2014) maintained that "I thus see no good reason for the claim that mathematics is a fundamentally different domain to science" (p. 38). If answers to Question 3 are positive, meaningful integration between mathematics and science should be enhanced. Otherwise, more autonomy can be granted to mathematics and science educators to emphasize subject characteristics.

Methods

When five plausible scores in mathematics are correlated with five plausible scores in science, it will generate 25 Pearson correlation coefficients (i.e., 5x5=25). Consequently, Type I error could be inflated from the repeated correlation computing. In addition, Borga (2001) pointed out,

> Ordinary correlation analysis is dependent on the coordinate system in which the variables are described. This means that even if there is a very strong linear relationship between two multidimensional signals, this relationship may not be visible in an ordinary correlation analysis if one coordinate system is used, while in another coordinate system this linear relationship would give a very high correlation. (p. 3)

Since TIMSS mathematics and science performance has separate scales, the issue of scale variation could undermine the result interpretation. Borga (2001) reviewed multivariate analyses, and found that "CCA [Canonical Correlation Analysis] finds the coordinate system that is optimal for correlation analysis" (p. 3). In the next section, CCA is singled out for a thorough review to address Question 1.

Question 2 is built on Question 1 to resolve an issue on result aggregation. Because correlation coefficients (r) are not additive, Garcia (2010) cautioned that "One cannot add raw r values to compute an arithmetic average an r" (p. 2). Therefore, a holistic approach needs to be taken to summarize plausible score correlations between mathematics and science. CCA can be employed to aggregate the correlation information across two groups of variables (Hardoon, Szedmak, & Shawe-Taylor, 2004). The variable grouping is examined in the next section to support CCA application on plausible score correlation.

After the correlation index construction, the seventh grade results from TIMSS 1995 are merged with the correlation findings across 35 countries to study the impact of between-subject correlation on science achievement (Question 3). As James (2014) suggested, "STEM integration in middle school mathematics and science classes may have a positive impact on mathematics amongst seventh grade students who are participating in the STEM classes versus those who are not participating in the STEM initiative" (p. 3). Hence, the grade choice is not only based on the data availability, but also considered the practical importance in STEM education.

Results

TIMSS data were properly accessed in this investigation to reconfirm an average of plausible scores in TIMSS reports (Beaton et al., 1996a; 1996b). For the baseline study in 1995, a total of 35 countries participated in TIMSS mathematics and science assessment at the seventh grade. Built on this database, one may feel tempting to use the average plausible score in each subject and run a correlation analysis between mathematics and science. However, that approach inadvertently ignores variability among plausible scores in each subject.

Alternatively, TIMSS researchers developed a JACKREGP program to use "achievement plausible values as the dependent variable" for regression analyses (Foy, Arora, & Stanco, 2013, p. 40). Statistics Canada (2003) also recommended "an SPSS macro called JACKREGPV.SPS that computes the average multiple correlation [R^2] between the specified plausible values and independent variables" (p. 162). The R^2 result could support configuration of the correlation coefficient (r) between independent and dependent variables [i.e., $r=\sqrt{R^2}$].

In using the SPSS macro, Brese, Jung, Mirazchiyski, Schulz, and Zuehlke (2011) noted that "It effectively performs five regression analyses – one for each plausible value – and aggregates the results" (p. 86). While five plausible scores from one subject can be entered in the SPSS macro as the dependent variables, the other set of plausible scores must be entered as an independent variable one at a time. Otherwise, a colinearity issue will occur when the independent variables are highly correlated on the same measurement construct of student performance.

Therefore, the literature reconfirmed an issue of inflating Type I error from the repeated applications of the SPSS macro for each pairs of plausible scores in mathematics and science. More importantly, correlation coefficients cannot be directly added and averaged for reporting. For instance,

$$\sqrt{R_1^2 + R_2^2 + R_3^2 + R_4^2 + R_5^2 + ...} \neq \sqrt{R_1^2} + \sqrt{R_2^2} + \sqrt{R_3^2} + \sqrt{R_4^2} + \sqrt{R_5^2} + ...$$

StatSoft (2000) reconfirmed that "Because the value of the correlation coefficient is not a linear function of the magnitude of the relation between the variables, correlation coefficients cannot simply be averaged" (p. 10). Due to the non-additive nature of correlation coefficients, a new method must be explored to support the result aggregation. Through an extensive review of the research literature, StatSoft (2015) concluded,

> Canonical Correlation is an additional procedure for assessing the relationship between variables. Specifically, this analysis allows us to investigate the relationship between two sets of variables. For example, an educational researcher may want to compute the (simultaneous) relationship between three measures of scholastic ability with five measures of success in school. (p. 1)

In this study, five plausible values in mathematics are grouped *as one set of variables* and five plausible values in science are treated *as another set of variables*. Thus, canonical correlation is an appropriate method for examining the relationship between two sets of variables. The analyses between variable groups automatically control Type I error for statistical reporting, and provide an answer for result aggregation in Questions 1 and 2.

Student performance in mathematics and science has been released in TIMSS report for international comparison (Beaton et al., 1996a; 1996b). After completing canonical correlation analyses, the results are merged with mathematics and science performance scores *at the country level* to address Question 3. The combined data set is attached in Appendix 1. Built on an assertion that science uses mathematics as a tool (Hellier, 2014), science achievement scores are treated as a dependent variable. The inclusion of canonical correlation as a predictor automatically assumes co-existence of mathematics achievement as an explanatory variable.

Grounded on these empirical evidences, the following SPSS syntax is used to construct a model of linear regression:

```
REGRESSION
/ORIGIN
/DEPENDENT Science
/METHOD=ENTER Math corr.
```

As a result, the SPSS printout below showed significant impacts on science performance across 35 countries that participated TIMSS seventh grade assessment.

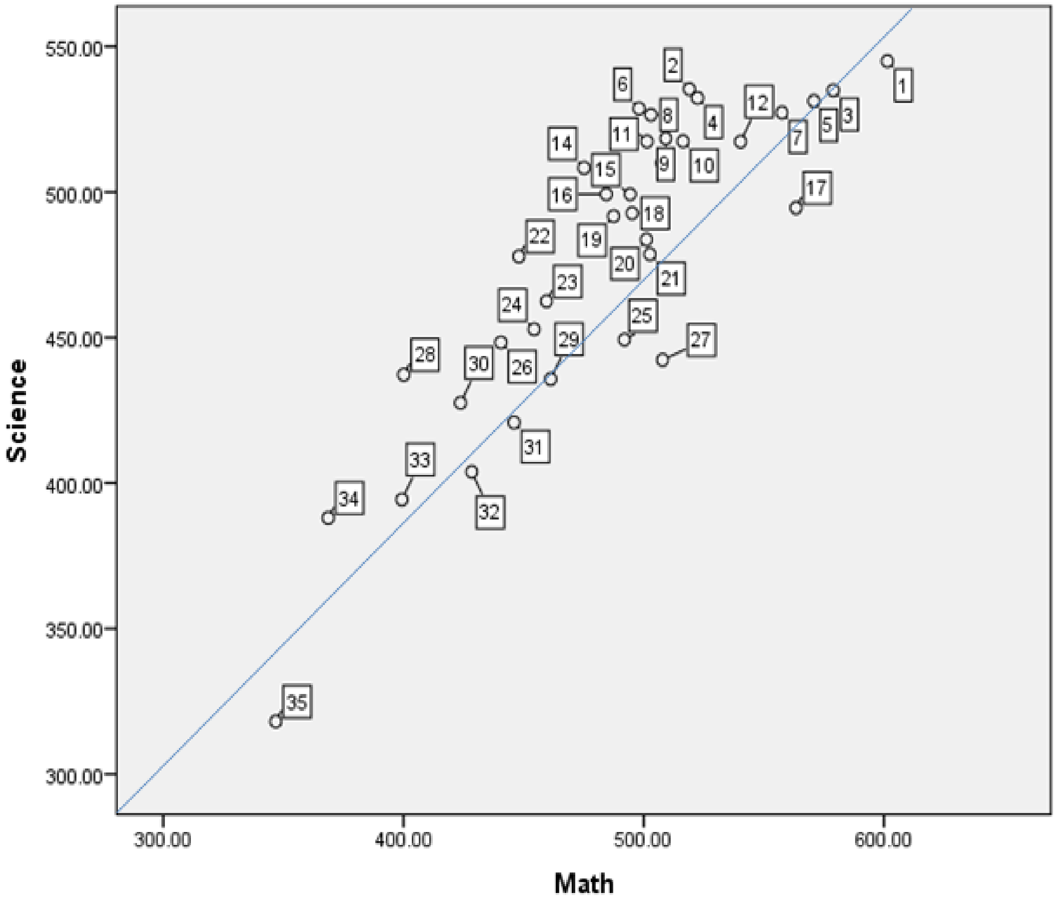
Table 1. Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	Math	.778	.070	.791	11.109	.000
	corr	159.372	54.274	.209	2.936	.006

a. Dependent Variable: Science
b. Linear Regression through the Origin

In addition, Figure 2 shows a scatter plot of science and mathematics scores from the TIMSS report results at the seventh grade. The results from 35 countries were sorted in a descending order so that Label "1" represents a country with the highest achievement in science and Label "35" indicates a country with the lowest science achievement. Figure 2 suggests that countries with a low rank in science tended to have low performance in mathematics. Hence incorporation of an indicator for the inter-subject correlation plays an important role in modeling the TIMSS outcomes in a cross-national context.

Figure 2. Scatter Plot of Mathematics and Science Scores



The mean and standard deviation for mathematics scores, science scores and their canonical correlations are listed in Table 2.

Table 2. Descriptive Statistics

Empirical Indicator	**N**	**Mean**	**Standard Deviation**
Science Performance Score	35	478.54	52.38
Mathematics Performance Score	35	485.87	57.41
Canonical Correlation Coefficient	35	.63	.06

Because performance scores were scaled in hundreds and correlation coefficient was between -1 and 1, the standardized regression coefficients provide a better index for comparing the impact on science performance from mathematics performance and the correlation between subjects. While mathematics achievement has β_1=.791, the subject correlation has β_2=.209 and both factors were significant at α=.01. As Srivastava and Ullah (1995) observed, "In applied work, the coefficient of determination (R^2) is most commonly used to judge the fit of a linear regression model" (p. 229). The adjusted coefficient of determination (R^2) reached .997, which reconfirmed a strong model fit for the comparative data from TIMSS.

Discussion

Along with rapid development of computer technology, statistical tools have become increasingly complicated and more useful in large data analyses (Larose, 2014). The complexity expands on both variable and subject dimensions. As each test item is coded as a variable in educational assessment, a lengthy test may involve a large number of variables. In this regard, the matrix-sampling technique has effectively addressed the item coverage for a lengthy test to match a "mile-wide" curriculum for international comparison. Therefore, Questions 1 and 2 of this investigation are grounded on the broad need of matrix sampling to examine plausible score correlation on the variable dimension.

Although TIMSS researchers developed a SPSS Macro syntax to average plausible score results, correlation coefficients are not additive. More importantly, Type I error could be inflated from repeated computations of correlation for each pair of plausible scores in mathematics and science. The issue was resolved in this study by introducing canonical correlation between two sets of plausible scores. In addition, the method can be implemented in simple SPSS application without involvement of complex Macro syntax on the variable dimension.

On the subject dimension, this method is robust for analyzing large-scale data that involve multistage sampling of schools and students. In the past, Kish (1965) employed *design effect* to describe increase of variability on statistical findings due to complex sampling. Wang and Ma (2006) further examined the impact on correlation computing. Because correlation coefficients (r) depend on a ratio of the *variance* and *covariance* components, the design effect impacts both numerator and denominator of the correlation computing. Thus, the influence from complex sampling is washed out.

While Figure 1 illustrate a positive correlation between mathematics and science scores across 35 countries, the standardized regression coefficients indicates that more impact on science performance from mathematics performance (β_1=.791) than the inter-disciplinary canonical correlation (β_2=.209) within each country. Thus, it is important to build on strong mathematics and science curricula for an inter-disciplinary integration. Because the reality is a unified whole, Hellier (2014) argued that knowledge of reality should be unified across different subject areas. His assertion was not only backed by empirical findings from this study, but also supported by past examples in history, such as Isaac Newton, who broke the grounds in mathematics and science concurrently.

Appendix 1 shows a range of canonical correlation from .51 in Belgium (Flemish Language) to .73 in Philippines. Neither country obtained higher mathematics or science scores than Singapore and Korea. Although an extreme emphasis or de-emphasis of inter-discipline connection might not be the best practice, it should be noted that South Africa (r=.57) and Columbia (r=.52) had one of the weakest canonical correlation between mathematics and science performance. Coincidently, they illustrated the lowest student performance in each discipline (Beaton et al., 1996a; b). Thus, the education quality is unlikely to reach a high level when a relatively low correlation is demonstrated between mathematics and science performance in the cross-national context.

In summary, this study is delimited to an analysis of the baseline TIMSS data at seventh grade across 35 countries. Before TIMSS, IEA's first and second international studies did not gather comparative data from the same group of students, nor did other projects, such as NAEP and PISA, fill this void. Although TIMSS created an unprecedented opportunity to examine correlation of mathematics and science achievement, it also introduced matrix sampling to generate five plausible scores in each discipline. This study was designed to disentangle methodological issues in aggregating correlational findings across plausible scores. As a result, Type I error was controlled through application of canonical correlation in each country. The comparative results reconfirmed significant relationship between mathematics and science performances. In addition, a low correlation in the interdisciplinary correlation tended to be linked to countries with a low performance scores in mathematics and science. These baseline findings are subjected to further verification by more comparative studies in a cross-national context.

References

Alphonse, L. (2014). *Talent pool coming up short*. Retrieved from http://bit.ly/1Uy FCPL.

American Educational Research Association (AERA) (2015). *AERA Research Grants Program*. Retrieved from http://bit.ly/1S4Wiwd.

Beaton, A., Martin, M., Mullis, I., Gonzalez, E., Smith, T., & Kelly, D. (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College, TIMSS International Study Center.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (1996b). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College, TIMSS International Study Center.

Borga, M. (2001). *Canonical correlation: A tutorial*. Retrieved from http://bit.ly/1X WgwB9.

Bradbum, N., Haertel, E., Schwille, J., & Tomey-Purta, J. (1991). A rejoinder to "I Never Promised You First Place". *Phi Delta Kappan*, *72*(10), 774-777.

Brese, F., Jung, M., Mirazchiyski, P., Schulz, W., & Zuehlke, O. (2011). *ICCS 2009 user guide for the international database*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement.

Brugger, K (2014). *Integrated math curriculum to replace traditional classes*. Retrieved from http://bit.ly/1Oosm3I.

Cogan, L. (1996, December). U.*S. eighth grade achievement in context. U.S. TIMSS National Research Center Report,* (pp. 1-5). East Lansing, MI: Michigan State University.

Davison, D. M., Miller, K. W., & Metheny, D. L. (1995). What does integration of science and mathematics really mean? *School Science and Mathematics, 95*, 226-230.

Fensterwald, J. (2014). *Districts split on high school math choices*. Retrieved from http://bit.ly/25X2bZF.

Foy, P., Arora, A., & Stanco, G. (2013). *TIMSS 2011 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Garcia, E. (2010). A tutorial on correlation coefficients. Retrieved from http://bit.ly/1QJlAVp.

Haigh, W., & Rehfeld, D. (1995). Integration of secondary mathematics and science methods courses: A model. *School Science and Mathematics, 95* (5), 240–247.

Hardoon, D., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation, 16*, 2639-2664.

Hellier, C. (2014). *Defending scientism: Mathematics is a part of science*. Retrieved from http://bit.ly/1sFVf1d.

International Association for the Evaluation of Educational Achievement (IEA) (2011). TIMSS 2011 Assessment Frameworks. Retrieved from http://bit.ly/1Yr5KBO.

Jacob, B., & Ludwig, J. (2009). Improving educational outcomes for poor children. In M. Cancian, & S. Danziger (Eds.), *Changing poverty, changing policies* (pp. 266–300). New York, NY: Sage.

Jakwerth, P. (1996). *Curriculum and achievement: Searching for the empirical link. U.S. TIMSS National Research Center Report* (pp. 10-12). East Lansing, MI: Michigan State University.

James, J. (2014). *Science, technology, engineering, and mathematics (STEM) curriculum and seventh grade mathematics and science achievement*. Phoenix, AZ: Grand Canyon University (UMI Number: 3614935).

Johnson, B. (2011). *How to creatively integrate science and math*? Retrieved from http://edut.to/1tq6hI8.

Johnson E. (1998). *Linking the national assessment of educational progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): A technical report*. Washington, DC: National Center for Education Statistics.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Larose, D. (2014). *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: John Wiley & Sons.

Lester, F. (2007). *Second handbook of research on mathematics teaching and learning*. Charlotte, NC: Information Age Publishing.

Lewis, S. P., Alacaci, C., O'Brrien, G. E., & Jiang, Z. (2002). Preservice elementary teachers' use of mathematics in a project-based science approach. *School Science and Mathematics, 102,* 172-180.

Martin, M. O., & Mullis, I.V.S. (2013). *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade – implications for early learning*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

McBride, J. W. & Silverman, F. L. (1991). Integrating elementary/middle school science and mathematics. *School Science and Mathematics, 91*(7), 18-25.

Mertens, S. B., Anfara, V. A, & Roney, K. (2009). *An international look at educating young adolescents*. Charlotte, NC: Information Age Publishing.

Mullis, I.V.S., & Martin, M. O. (2013). *TIMSS 2015 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Science Foundation (NSF) (2013). *Inspiring STEM learning*. Retrieved from http://1.usa.gov/1XoKcpL.

Rotberg, I. (1990). I never promised you in first place. *Phi Delta Kappan, 72* (4), 296-303.

Rotberg, I. (1991). How did all those dumb kids make all those smart bombs? *Phi Delta Kappan, 72* (10), 778-781.

Rotberg, I. (1995). Myths about test score comparisons. *Science, 270*, 1446-1448.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass Publishing.

Srivastava, A., & Ullah, A. (1995). The coefficient of determination and its adjusted version in linear regression models. *Econometric Reviews, 14* (2), 229-240.

Statistics Canada (2003). *The International Adult Literacy and Skills Survey, 2003*. Retrieved from http://bit.ly/1sFVLw9.

StatSoft (2000). *Basic statistics*. Retrieved from http://bit.ly/1rpeZFf.

StatSoft (2015). *How to assess the relationship between variables, canonical analysis*. Retrieved from http://bit.ly/1sFWWvH.

Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal, 32*, 227-249.

Vygotsky, L. (1987). *Thought and language*. Cambridge, MA: MIT Press.

Wang, J. (1998). International achievement comparison: Interesting debates on inconclusive findings. School Science and Mathematics, *98*(7), 376-382.

Wang, J. (2011). Re-examining test item issues in the TIMSS mathematics and science assessments. *School Science and Mathematics, 111* (7), 334-344.

Wang, J. & Ma, X. (2006, April). *A study of correlation between the NAEP scores and student coursework in science*. Paper accepted for presentation at the 2006 annual meeting of the American Educational Research Association (AERA), San Francisco, CA.

Will, M. (2014). *In transition to Common Core, some high schools turn to 'Integrated' Math*. Retrieved from http://bit.ly/1ssdrDK.