Relationship Between Mathematics and Science Achievement at the 8th Grade

Jianjun Wang
Department of Advanced Educational Studies
California State University, Bakersfield
CA 93311

ABSTRACT

Mathematics and science achievements have been assessed in the Third International Mathematics and Science Study (TIMSS) and its repetition (TIMSS-R). Most existing reports are largely separated into the subject domains. This article is focused on an examination of the relationship between mathematics and science achievements based on student test scores. Moderate correlation coefficients have been found from the statistical analyses on the TIMSS and TIMSS-R measurement scales. These empirical findings may help mathematics and science educators assess the need for curriculum integration advocated by several professional organizations in the U.S. and other nations.

Introduction

The global market competition has been one of the driving forces for enhancing educational accountability in many countries (Martin et al., 1998; 2001). As a result, professional guidelines have been developed over the last decade to strengthen mathematics and science curriculum standards. In the United States, national organizations produced documents to advocate curriculum articulation between mathematics and science education (e.g., National Council for Teachers of Mathematics, 2000; National Research Council, 1996). Meanwhile, educators in the United Kingdom adopted interdisciplinary approaches in the development of its national curriculum (Nixon, 1991). The Curriculum Council of Western Australia (1998) also recommended teaching methods across traditional subject boundaries (Venville, Wallace, Rennie, & Malone, 1998). Implementation of these initiatives around the world ranges from thematic units to an entirely combined curriculum (Lonning, DeFranco, & Weinland, 1998). According to Haigh and Rehfeld (1995), "most of these attempts have been based upon the assumption that integration increases student achievement in both mathematics and science" (p. 241).

Widely cited as an international benchmark, student achievement scores were released from the *Third International Mathematics and Science Study* (TIMSS) in 1995 and a repeat of the TIMSS project (TIMSS-R) in 1999. These projects incorporated both mathematics and science tests (e.g., Martin & Mullis, 1996; Mullis, et al., 2000). Therefore, correlation coefficients between mathematics and science scores can be analyzed to assess the intersubject relationship at the eighth grade.

In terms of the content structure, the relationship seems to be asymmetric between mathematics and science. "Unlike the mathematics teacher who can choose to avoid science, the science teacher is not able to cover most topics without calling on mathematical concepts and skills" (Frykholm & Meyer, 2002, p. 504). Furthermore, the reliance on mathematics varies across the different science fields. Among the courses taught in most schools, physics is a subject heavily dependent on mathematical skills (Basson, 2002; Hanna & Jahnke, 2002; Meltzer, 2002). However, the demand on mathematical preparation is not as strong in biology, and "other sciences such as psychology might not yet be ready for the kind of mathematization that has taken place in physics" (Orton & Roper, 2000, p. 124).

In comparative education, TIMSS test scores represent "those parts of the intended and implemented curriculum which are actually learned by the students" (International Bureau of Education, 2002, p. 1). Whereas it has been assumed that "integration would produce greater learning outcomes of both mathematics and science, … few empirical attempts have been made to test this assumption" (McBride & Silverman, 1991, p. 286). To date, the TIMSS and TIMSS-R reports are divided into mathematics and science domains (e.g., Beaton et al., 1996a, b; Martin et al., 1998, 2001; Mullis, et al., 1998, 2001), and no results have been disseminated on the relationship of student performance between these subjects. In this regard, this investigation may enrich the existing research literature by adding new empirical findings across the subject boundaries.

Review of the Literature

Lederman and Niess (1998) observed that "the current reforms have resulted in renewed interest in curriculum integration, especially between mathematics and science" (p. 281). Given the emphasis on the quality of mathematics and science education (e.g., NCTM, 1998; Nixon, 1991; NRC, 1996; Venville et al., 1998), more interdisciplinary research needs to be conducted at the national or international levels to analytically address two fundamental topics: (1) a system-wide assessment of correlation between mathematics and science achievements; and (2) an examination of the linkage between a higher correlation and a higher average score in mathematics or science (Czerniak, Weber, Sandmann, & Ahern, 1999; Hurley, 2001; Pang & Good, 2000).

In the United States, the *National Assessment of Educational Progress* (NAEP) has been one of the primary measures used to assess the condition of education for more than three decades. Special research designs from the NAEP survey are useful to empirical studies of student performance in comparative education.

To help understand the details of the NAEP contribution, the data collection process can be described in terms of teachers' daily teaching practice. In a typical classroom setting, science or mathematics teachers often face a challenge of squeezing much of the content into a final examination at the end of a semester. For a large-scale study, a more extensive test is needed to assess student cumulative performance over multiple grades from different nations.

A lengthy test may interfere with various psychological constructs, such as student speed, fatigue, and mental stability, in the testing process. Because the score differences could be interpreted by those factors other than academic achievement, the lengthy testing may undermine the internal validity of the performance scores (Best & Khan, 2003). According to the joint technical standards of educational and psychological testing (AERA, APA, & NCME, 1999), the reliability of test scores is defined by the measurement consistency. When multiple constructs are involved, no single index can be used to assess the reliability of the measurement outcome (Hambleton & Swaminathan, 1985).

A solution developed by the NAEP researchers is to split the lengthy test into multiple components, and group those components into several concise test booklets. Each student is tested on one booklet that covers a portion of the lengthy instrument. A balanced coverage of the entire instrument can be achieved when all components of the test are randomly taken by an equal number of students from each education system. In addition, based on the performance from part of the instrument, total scores can be obtained at the individual level from statistical imputation (Pashley & Phillips, 1993). To assess stability of the "manufactured" data, a total of five plausible scores have been computed in NAEP to represent student performance (Allen, Carlson, & Zelenak, 1999).

Before introducing the NAEP methodology, student test burden was a long-lasting problem in the first and second international studies of mathematics or science achievement in 1964 and in 1985. TIMSS is the first large-scale comparative project that incorporates the plausible value computation. More importantly, TIMSS and its repetition (TIMSS-R) are designed to test student performance in both mathematics and science subjects. This linkage has laid a solid cornerstone for a correlation analysis of student scores.

The released test scores also represent an articulated effort in the research design. After TIMSS-R, the original TIMSS scores were rescaled to enhance comparability between these two databases (Martin, Gregory, & Stemler, 2000). Therefore, both TIMSS and TIMSS-R databases can be employed to examine the score correlation on different measurement scales to assess consistency of the research findings between these projects.

Most statistical computations are grounded on an assumption of simple random sampling. In TIMSS and TIMSS-R, the data were collected from a stratified sample to build on characteristics of each education system. Consequently, the variance of the research findings is typically larger than the corresponding variance from a simple random sample (Martin, Gregory, & Stemler, 2000; Martin & Kelly, 1997). Kish (1965) introduced a concept of design effect (deff) to describe the difference:

$$\textbf{deff} = \frac{\textbf{variance from stratified sampling}}{\textbf{variance from random sampling}}$$

Because of the larger statistical variability associated with the stratified sampling, one may expect a deff value larger than 1. The exact deff value, however, varies among specific variables. For a large-scale data analysis, special software packages are needed to handle the statistical computing from stratified sampling (Cabrera, La Nasa, & Burkum, 2002). The consideration of statistical variability is important to this correlation study because correlation coefficients depend on the variance indexes (Ott, 1993):

$$r = \frac{cor(x_1, x_2)}{\sqrt{\operatorname{var}(x_1)\operatorname{var}(x_2)}}$$

One of the widely used software packages for this type of analyses is an AM program developed by the American Institute of Research (AIR). AIR (2003) noted,

> AM is a statistical software package for analyzing data from complex samples, especially large-scale assessments such as the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Studies (TIMSS) (http://am.air.org, p. 1).

In summary, TIMSS researchers have adopted the modern assessment methodology to effectively reduce student testing burden. Due to statistical uncertainty in the achievement assessment, five plausible scores have been imputed in each subject area, and "one set of the imputed plausible scores can be considered as good as another" (Gonzalez & Smith, 1997, ch. 6, p. 3). This interchangeability also suggests equivalency of the design effect (deff) among the plausible scores from the stratified sample design. Under an assumption of the invariant deff values between mathematics and science scores, the AM software is adopted to compute correlation coefficients from the TIMSS and TIMSS-R data sets gathered through stratified sample designs.

Research Questions

On the basis of the international data downloadable at http://www.timss.org, research questions that guide this investigation are:

1. Is there a linear relationship between student mathematics and science achievements from TIMSS/TIMSS-R participating countries?

2. Is the higher correlation linked to a higher average *mathematics* or *science* performance in the international comparison?

3. What are the consistent findings from the resulting triangulation across TIMSS and TIMSS-R projects?

Methods

At an initial stage of this data analysis, the average mathematics and science scores have been computed to reconfirm results from the existing TIMSS/TIMSS-R reports (Beaton et al., 1996a, b; Martin, et al., 2000; Mullis, et al., 2000). To verify a linear relationship between mathematics and science achievement, plots are created to examine the score patterns across the participating countries (Question 1). Because of the involvement of multiple plausible scores, a summary index is needed to average the coefficients from the correlation matrix. According to Corey, Dunlap, and Burke (1998), "When correlations come from a matrix, there is a consistent advantage associated with using [Fisher's] z'. Across sample size and numbers of correlations averaged, bias in average r(z)' is smaller than bias in average r" (p. 260). To reduce the statistical bias, the Fisher's (1921) z transformation is conducted for each nation to compute an average correlation coefficient among plausible scores in mathematics and science.

By viewing the world as a giant education laboratory, the diverse education settings in each country may have resulted in different correlation coefficients between students' mathematics and science scores. To facilitate the system-wide comparison, the average correlation coefficient is examined at the country level to determine if a higher coefficient corresponds to a higher average score in mathematics or science (Question 2).

Patterns of the data distribution are plotted to verify consistency of research findings between TIMSS and TIMSS-R databases on the old and new scales (Question 3). Results of the correlation analysis are articulated with TIMSS/TIMSS-R scores to discuss educational implications of the interdisciplinary emphasis advocated by several professional organizations.

Results

Mathematics and science scores for all participating countries have been plotted on three scales: (1) TIMSS results on the TIMSS original scale; (2) TIMSS results on the new TIMSS-R three-parameter (difficulty, discrimination, and guessing) scale; and (3) TIMSS-R results on the new scale (see Figures 1-3).

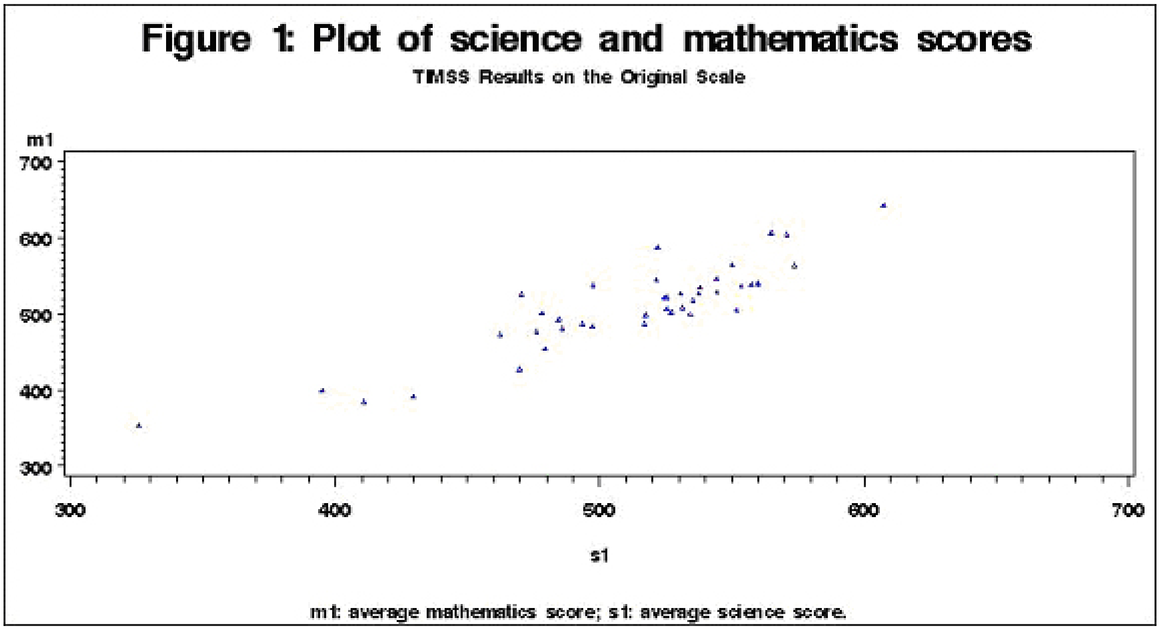


Figure 1: Plot of science and mathematics scores
TIMSS Results on the Original Scale
m1
700
600
500
400
300
300
400
500
600
700
s1
m1: average mathematics score; s1: average science score.

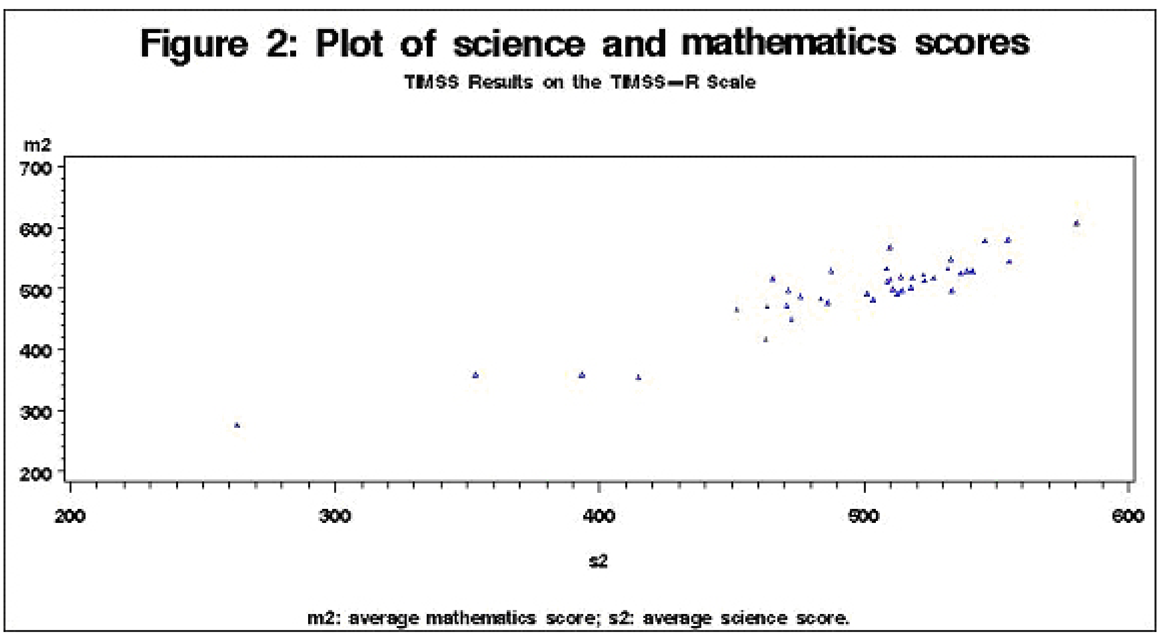
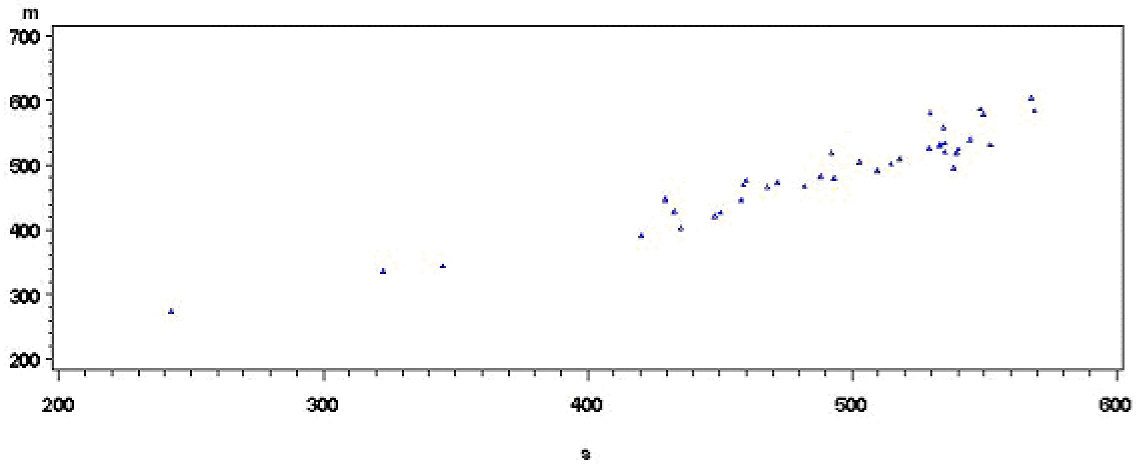


Figure 2: Plot of science and mathematics scores
TIMSS Results on the TIMSS-R Scale
m2
700
600
500
400
300
200
200
300
400
500
600
s2
m2: average mathematics score; s2: average science score.

Figure 3: Plot of mathematics and science scores

TIMSS—R Results



m: average mathematics score; s: average science score.

The average correlation coefficients from the Fisher's z transformation are listed in Table 1 for each nation.

Table 1. Correlation coefficients between mathematics and science scores

Country	r_95	r_new95	r_99
Australia	0.61066	0.72561	0.73445
Austria	0.60400	0.74650	.
Belgium (FL)	0.53763	0.72533	0.76603
Belgium (FR)	0.56130	0.72329	0.67450
Bulgaria	.	.	0.70826
Canada	0.50309	0.60844	0.65022
Chile	.	.	0.67055
Chinese Taipei	.	.	0.79888
Colombia	0.46660	0.70264	.
Cyprus	0.60592	0.72942	0.71988
Czech	0.55550	0.71477	0.68582
Slovak	0.59682	0.72885	.
Denmark	0.54204	0.67358	.
Finland	.	.	0.63206
France	0.44642	0.59251	.
Germany	0.60900	0.76270	.
Greece	0.57560	0.71952	.
Hong Kong	0.55414	0.72474	0.69972
Hungary	0.58429	0.70712	0.71063
Iceland	0.52166	0.69369	.

(continued)

Table 1. Correlation coefficients between mathematics and science scores

Country	r_95	r_new95	r_99
Indonesia	.	.	0.69388
Iran	0.43267	0.62444	0.66955
Ireland	0.59635	0.73103	.
Israel	0.60736	0.71732	0.77825
Italy	.	.	0.76616
Japan	0.55408	0.68190	0.72506
Jordan	.	.	0.75202
Korea	0.57265	0.72091	0.73592
Kuwait	0.38991	0.55434	.
Latvia	0.50811	0.64808	0.67064
Lithuania	0.54685	0.65489	0.73885
Malaysia	.	.	0.72142
Moldova	.	.	0.68617
Morocco	.	.	0.43424
Netherlands	0.56049	0.75407	0.69678
New Zealand	0.58807	0.72179	0.76637
Norway	0.54790	0.66924	.
Philippines	0.62747	0.79023	0.66908
Portugal	0.45528	0.64280	.
Romania	0.60999	0.71681	0.70437
Russian	0.56386	0.67507	0.71900
Singapore	0.49319	0.68224	0.78195
Slovenia	0.55386	0.70312	0.70417
South Africa	0.54361	0.89345	0.67490
Spain	0.48891	0.64371	.
Sweden	0.54953	0.69207	.
Switzerland	0.55418	0.73575	.
Thailand	0.51430	0.61442	0.71423
Tunisia	.	.	0.54241
Turkey	.	.	0.65385
Macedonia	.	.	0.71579
England	0.60866	0.72191	.
Scotland	0.58280	0.71281	.
USA	0.61213	0.74919	0.77763

Notes: (1) Belgium (FL, FR), Czech, Slovenia, and South Africa have been assigned different country IDs in the TIMSS and TIMSS-R data codebooks.

(2) For those countries that did not participate in both TIMSS and TIMSS-R, the sign "." is the default for missing observations

(3) r_95 : Correlation coefficients from the original TIMSS scale;

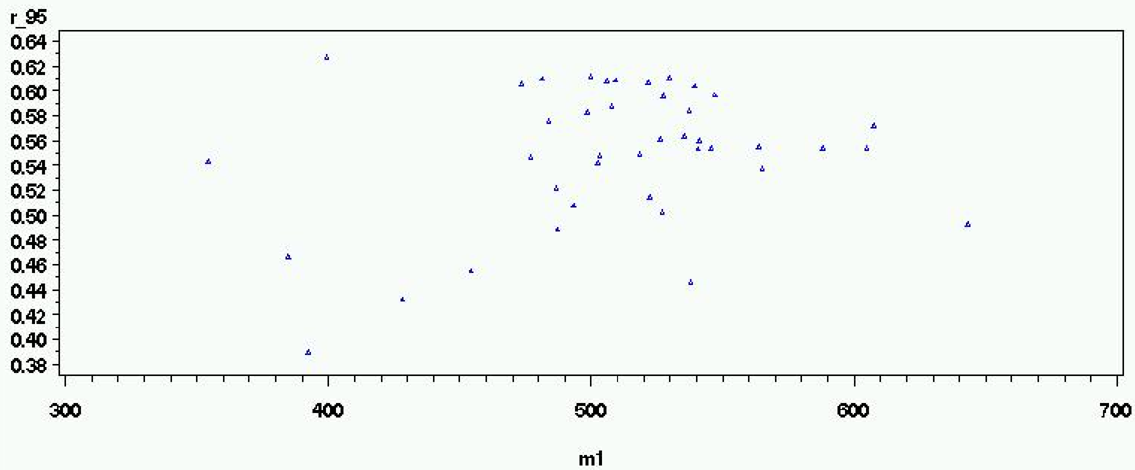
r_new: TIMSS correlation coefficients on the new TIMSS-R scale;

r_99 : Correlation coefficients from the TIMSS-R database.

Because student scores represent an achieved curriculum in an education system (Linn, 2000; Zabulionis, 2001), the correlation coefficient indicates relationship of the achieved curricula between mathematics and science. The correlation coefficient has been plotted to show its relationships with academic performances in mathematics (Figures 4-6) and science (Figures 7-9).

Figure 4: Plot of correlation coefficients and math scores

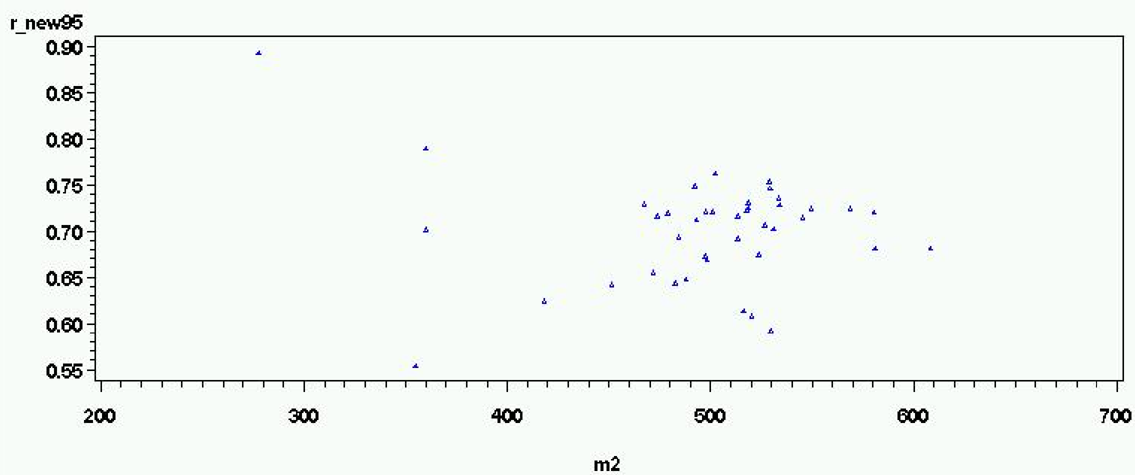
TIMSS Results on the Original Scale



r_95: correlation coefficient; m1: average mathematics score.

Figure 5: Plot of correlation coefficients and math scores

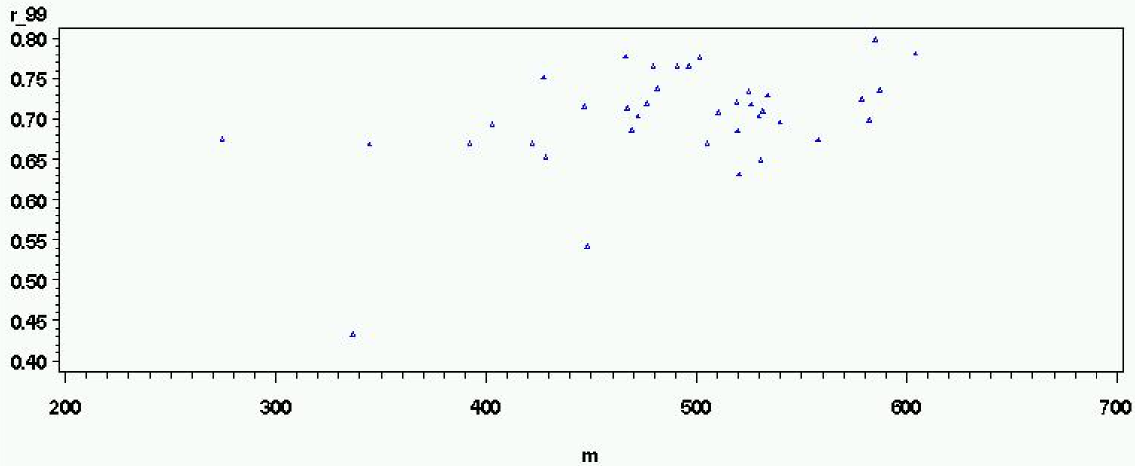
TIMSS Results on the TIMSS–R Scale



r_new95: correlation coefficient; m2: average mathematics score.

Figure 6: Plot of correlation coefficients and math scores

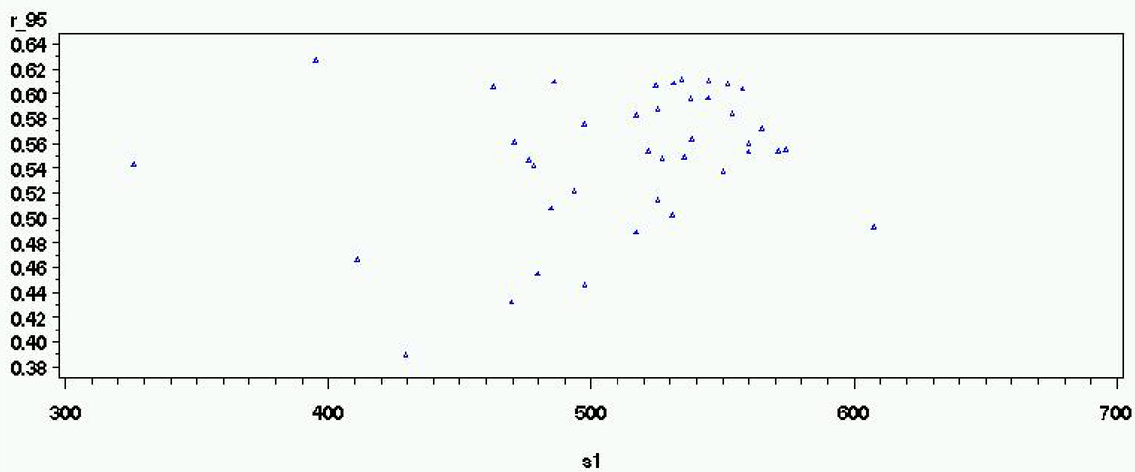
TIMSS–R Results



r_99: correlation coefficient; m: average mathematics score.

Figure 7: Plot of correlation coefficients and science scores

TIMSS Results on the Original Scale



r_95: correlation coefficient; s1: average science score.

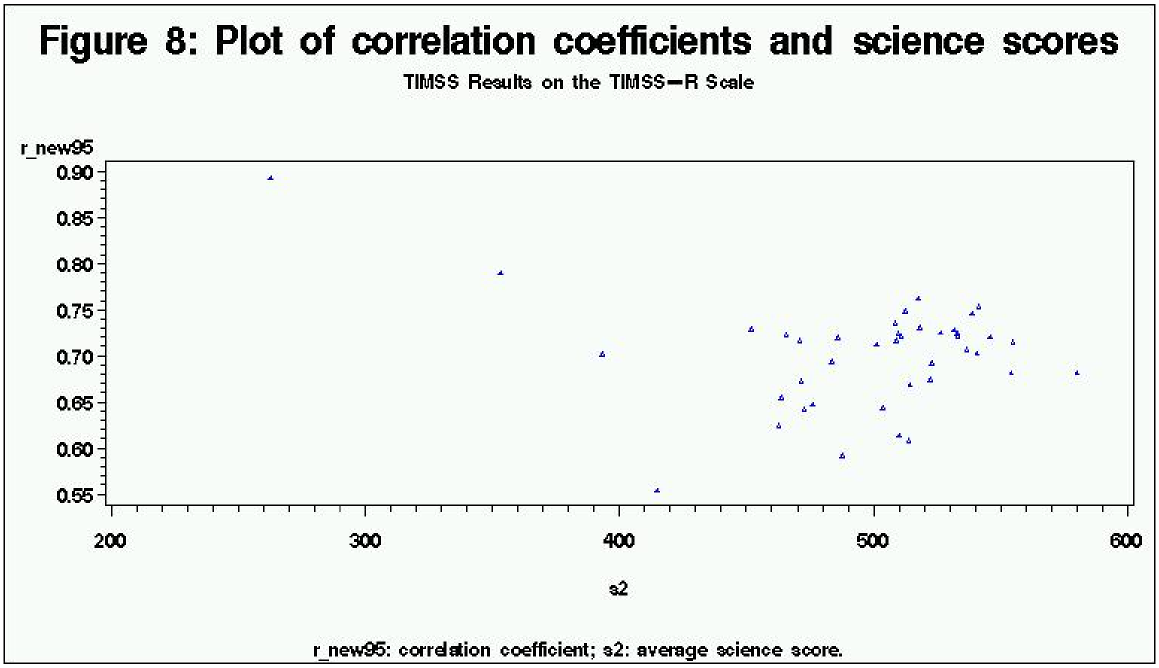


Figure 8: Plot of correlation coefficients and science scores
TIMSS Results on the TIMSS–R Scale
r_new95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
200
300
400
500
600
s2
r_new95: correlation coefficient; s2: average science score.

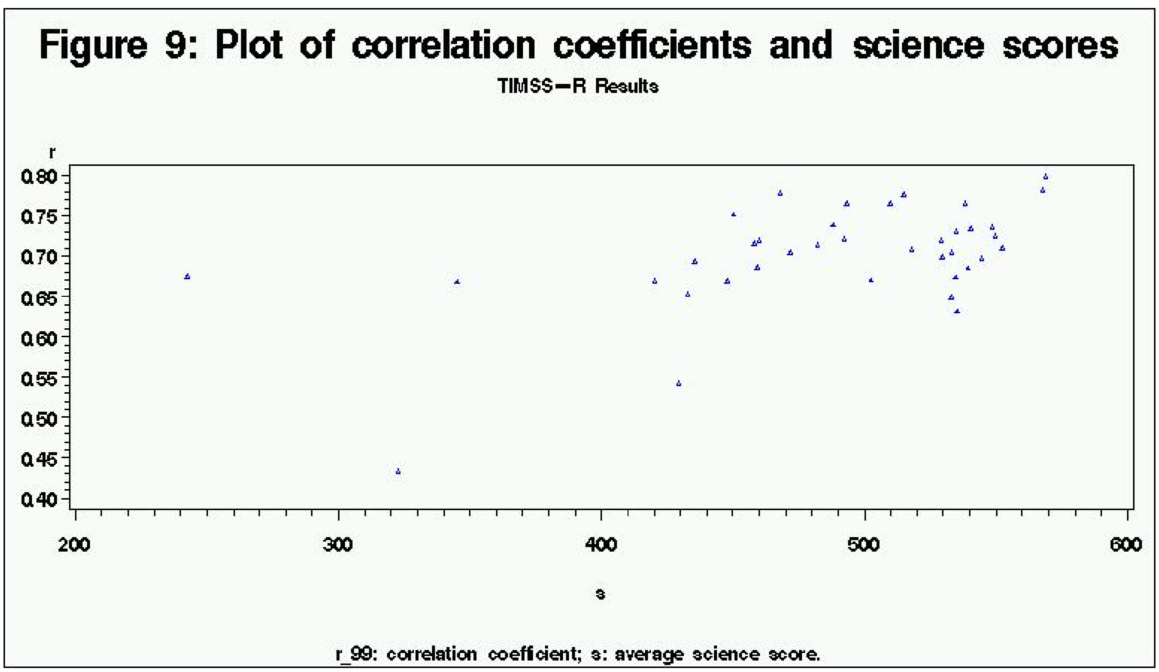


Figure 9: Plot of correlation coefficients and science scores
TIMSS–R Results
r
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
200
300
400
500
600
s
r_99: correlation coefficient; s: average science score.

The correlation coefficients have been computed to describe relationships between the curriculum link (r) and student scores on both new and old scales (Table 2).

Table 2. Correlation between student performance and the indicator of math-science link

	r_95	r_new95	r_99
Mathematics achievement	0.24960	-0.15469	0.46436 ***
Science achievement	0.25681	-0.25216	0.51684 ***

Notes: (1) The math-science link in the achieved curriculum is described by the correlation between mathematics and science achievements.
(2) *** indicate that the correlation is significant at 0.05 level.
(3) r_95 : Correlation coefficients from the original TIMSS scale;
r_new: TIMSS correlation coefficients on the new TIMSS-R scale;
r_99 : Correlation coefficients from the TIMSS-R database.

Discussion

Relations between mathematics and science education have been established in disciplinary roots of each subject. "Science provides rich contexts and concrete phenomena demonstrating mathematical patterns and relationships. Mathematics provides the language and tools necessary for deeper analysis of science concepts and applications" (Basista & Mathews, 2002, p. 359). The existing knowledge has been largely confined within the qualitative description, and few quantitative indexes have been developed from large-scale databases (McBride & Silverman, 1991). In part, this was because domestic projects like NAEP did not gather mathematics and science performance from the same group of students. Thus, no correlation analysis can be conducted to disentangle the interdisciplinary linkage between these two core subjects.

Fortunately, TIMSS and TIMSS-R have incorporated the mathematics and science testing, and correlation coefficients between scores in these subjects have been computed for the participating nations (Table 1). Whereas the international curricula may vary in its objectives and contents, student scores on the same measurement scale are directly comparable, and represent the education outcomes demonstrated by the student performance. Accordingly, results of the statistical investigation are discussed following the sequence of the research questions that guided the TIMSS and TIMSS-R data analyses.

Confirmation of the Linear Relationship

Variables could be linked in a variety of different ways. Linear relationship is the simplest pattern to describe concurrent changes between any two variables. Under this linear condition, variable relations can be summarized by a single correlation coefficient (Ott, 1993), which substantially simplifies relevant interpretations of the research outcomes. In graphic representations, the linear relation is illustrated by a straight line and delineates the shortest distance between two points. Otherwise, an infinite number of curves can be used to link those points. Consequently, when one variable changes, there is an infinite number of ways to project the associated changes on the other variable.

Clearly, ascertaining the relationship pattern is a crucial step for the correlation analysis. Figures 1-3 unanimously confirm the linear model in scatterplots of mathematics and science average scores from the TIMSS and TIMSS-R results. Besides the pattern match, Table 1 also shows that the linear correlation coefficients are informative enough to differentiate the linkage between mathematics and science achievement among different countries. Built on these observations, the linear correlation coefficient can be employed as an appropriate indicator to link the result of student mathematics and science education.

Linkage of the Correlation with Academic Performance

Like the existence of various levels of curriculum integration between mathematics and science, the linkage between mathematics and science scores also shows a fair amount of variability, ranging from Kuwait's correlation coefficient of 0.39 on the TIMSS original scale (country id: 414) to South Africa's 0.89 from the rescaled scores in TIMSS-R (country id: 717) (Table 1). These relationship indicators can be articulated in an assessment of education quality in mathematics and science. Lehman (1994) asserted that one of the reasons for mathematics and science integration was "to increase students' achievement in both disciplines" (p. 58).

Regardless of the subject domains or research projects, Figures 4-9 show that the test scores may vary from a level below 300 points to a peak above 600 points. For countries with a performance score below 450, the range of the correlation coefficients is much larger than the range of top-performing countries scoring around 600 points (Figures 4-9). More specifically, none of the top nations have the highest correlation of .89 nor the lowest correlation of .39. Regardless of the country identity or the curriculum setting, the results seem to support mathematics and science instructions that lead to a moderate correlation of student achievement scores between these two subjects.

In retrospect, when mathematics and science were taught as two separate subjects, professional organizations have encouraged interdisciplinary integration to develop more coherent school education (e.g., National Council for Teachers of Mathematics, 2000; National Research Council, 1996). It would be far fetched to simply interpret this encouragement as "the more the integration, the better the curriculum." Underhill (1995) pointed out, "There is a substantial body of mathematical knowledge which has no direct relation to science (e.g., theoretical mathematics). There is a substantial body of scientific knowledge which has no direct relation to mathematics (e.g., the names of things and their parts)" (p. 225). Thus, a moderate correlation appears to match the partial overlap between these subjects.

In short, based on this investigation, a more balanced position should be taken by school professionals to avoid an overemphasis or de-emphasis of instructional integration between mathematics and science subjects. Across all participating nations around the world, the average score correlation accounts for no more than 22% (i.e., $r^2 < .22$ in Table 2) of mathematics achievement and no more than 27% (i.e., $r^2 < .27$ in Table 2) of science achievement.

Triangulation of the Results Across TIMSS and TIMSS-R

Whereas the aforementioned results are invariant, the range of correlation coefficient does vary among the measurement scales (Table 1). In the original TIMSS scale, only item difficulty was modeled in terms of the student response. TIMSS-R adopted a three-parameter model to estimate parameters of difficulty, discrimination, and guessing. The scale difference has been reflected in the correlation outcomes. Under the one-parameter model, the correlation of student achievements in TIMSS appears to be positively associated with average national performance in mathematics and science (Table 2). But when the TIMSS data are rescaled on the three-parameter model, the correlation becomes negative (Table 2). Figures 4, 5, 7, and 8 show a slight upward or downward pattern from the results. A close examination of Figures 5 and 8 also reveals an outlier with low performance scores and a high correlation coefficient between mathematics and science achievements (r = 0.89) from South Africa. Had this observation been taken out, the correlation coefficients on the new scale would be 0.19 and 0.14, instead of -0.15 and -0.25, respectively (Table 2). Therefore, the negative correlation might reflect statistical artifact, and the scale transformation has an impact on the correlation findings.

In contrast, the TIMSS-R results of 1999 show a significant correlation between correlation coefficient values and student achievement in mathematics or science (Table 2). It should be noted that only "twenty-six countries took part in the TIMSS eighth-grade assessments in both 1995 and 1999" (Mullis, et al., 2000, p. 34). Twelve countries participated in only one of the international studies. Given the difference in participating nations, the correlation results could have been affected by the involvement of different countries between TIMSS and TIMSS-R.

The advantage of the three-parameter scale over the one-parameter scale largely hinges on whether it is necessary to model the effect of guessing (Hambleton & Swaminathan, 1985). Hambleton (1988) noted, "with difficult multiple-choice tests, a researcher might anticipate considerable guessing on the part of examinees. Needed, therefore, would be a model that could handle this situation" (p. 154).

Because more than 90% of the TIMSS items are in a multiple-choice format (Lange, 1997), the result seems to support the effort of TIMSS researchers in rescaling the TIMSS results, taking into consideration the guessing effect (Martin et al., 2000; Mullis et al., 2000). For instance, the TIMSS instrument has the following mathematics item in a multiple-choice format:

O2. If the price of a can of beans is raised from 60 cents to 75 cents, what is the percent increase in the price?

A. 15%

B. 20%

C. 25%

D. 30%

Figuring out the *rate increase* is a basic mathematical skill required in many scientific experiments. With the four options in this question, the probability of obtaining a correct answer through random guessing is 25%. In the TIMSS data, only 28% of 8th graders from all participating nations answered this question correctly! This result not only illustrates a need for correcting the guessing effect, but also urges educators to make a concerted effort to improve student performance in this joint area between mathematics and science.

In summary, this empirical data analysis seems to suggest that the call for subject articulation is still a valid accountability issue following the long-lasting quest for educational improvement. However, more attention should be given to the research methodology that is directly linked to validity of the empirical data under investigation. Without recommending a particular mode of curriculum integration, this research is focused on correlation of student achievement between mathematics and science on a common international scale. While educators may continue to exercise their creativity in developing more effective curricula, this research seems to support instructional initiatives that result in a moderate correlation in student scores between mathematics and science.

About the Author

Jianjun Wang is a full professor of educational research and statistics at California State University, Bakersfield. He is interested in stochastic processes and comparative studies in mathematics and science education.

Correspondence concerning this article should be addressed to Jianjun Wang, Department of Advanced Educational Studies, California State University, Bakersfield, 9001 Stockdale Highway, Bakersfield, California, CA 93311. Home: (661) 664-3048. Fax (661) 664-2016. E-mail: jwang@csub.edu.

References

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (1999). *The standards for educational and psychological testing.* Washington, DC: AERA.

American Institute of Research [AIR] (2003). AM statistical software. Retrieved March 26, 2003 from http://am.air.org .

Basista, B. & Mathews, S. (2002). Integrated Science and Mathematics Professional Development Programs. *School Science and Mathematics, 102* (7), 359-370.

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *International Journal of Mathematical Education in Science and Technology, 33* (5), 679-690.

Beaton, A., Martin, M., Mullis, I., Gonzalez, E., Smith, T., & Kelly, D. (1996a). *Science achievement in the middle school years.* Chestnut Hill, MA: Boston College.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (1996b). *Mathematics achievement in the middle school years.* Chestnut Hill, MA: Boston College.

Best, J. & Kahn, J. (2003). *Research in education* (9th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Cabrera, A. F., La Nasa, S. M., & Burkum, K. R. (2002). Pathways to college: What_affects lowest-SES students' decisions to transfer to a four-year institution? Retrieved December 20, 2004 from http://www.eriche.org/ASHEpapers02/Cabrera%20-%20Pathways%20to%20College.pdf.

Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Observed and expected bias in average correlation with and without using Fisher's z transformation. *Journal of General Psychology, 125*, 245-261.

Curriculum Council of Western Australia (1998). Curriculum framework consultation draft. Perth, Western Australia: Author.

Czerniak, C., Weber, W., Sandmann, A., & Ahern, J. (1999). A literature review of science and mathematics integration. *School Science and Mathematics, 99* (8), 421-430.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1, 1-32.

Frykholm, J. A. & Meyer, M. R. (2002, May). Integrated instruction: Is it science? Is it mathematics? *Mathematics Teaching in the Middle School*, 502-508.

Gonzalez, E. J., & Smith, T. A. (1997). *Users guide for the TIMSS international database.* Chestnut Hill, MA: TIMSS International Study Center.

Haigh, W. & Rehfeld, D. (1995). Integration of secondary mathematics and science methods courses: A model. *School Science and Mathematics, 95* (5), 240-244.

Hambleton, R. K. (1988). Principles and selected applications of item response theory. In R. L. Linn (Ed), *Educational measurement* (3rd ed.). London: Collier Macmillan.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston, MA: Kluwer.

Hanna, G. & Jahnke, H. (2002). Arguments from physics in mathematical proofs: An educational perspective. *For the Learning of Mathematics, 22* (3), 38-45.

Hurley, M. (2001). Reviewing integrated science and mathematics: The search for evidence and definitions from new perspectives. *School Science and Mathematics, 101* (5), 259-268.

International Bureau of Education (2002). Attained curriculum. Retrieved September 8, 2003 from http://www.ibe.unesco.org/International/DocServices/Thesaurus/ 00003341.htm.

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

Lange, J. D. (1997). Looking through the TIMSS mirror from a teaching angle. Retrieved October 11, 2000 from http://www.enc.org/topics/timss/additional/documents/ 0,1341,CDS-000158-cd158,00.shtm.

Lederman, N. G. & Niess, M. L. (1998). 5 apples + 4 oranges = ? *School Science and Mathematics, 98* (6), 281-284.

Lehman, J. R. (1994). Integrating science and mathematics: Perceptions of preservice and practicing elementary teachers. *School Science and Mathematics, 94*, 58-64.

Linn, R. (2000). The measurement of student achievement in international studies. Paper prepared for the Board for International Comparative Studies in Education. Retrieved December 22, 2004 from http://www4.nas.edu/DBASSE/BICSE.nsf/files/ Linn.pdf/).

Lonning, R., DeFranco, T., & Weinland, T. (1998). Development of theme-based, interdisciplinary, integrated curriculum: A theoretical model. *School Science and Mathematics, 98* (6), 312-318.

Martin, M. O., Gregory, K. D., & Stemler, S. E. (2000). *TIMSS 1999: Technical report*. Chestnut Hill, MA: TIMSS International Study Center.

Martin, M. O. & Kelly, D. L. (1997). *Technical report volume II: Implementation and analysis.* Chestnut Hill, MA: TIMSS International Study Center.

Martin, M. O. & Mullis, I. V. S. (1996). *Quality assurance in data collection*. Chestnut Hill, MA: TIMSS International Study Center.

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1998). *Science achievement in Missouri and Oregon in an international context: 1997 TIMSS benchmarking*. Chestnut Hill, MA: TIMSS International Study Center.

Martin, M. O., Mullis, I., Gonzalez, E., O'Connor, Chrostowski, S., Gregory, K., Smith, T., & Garden, R. (2000) *TIMSS 1999: International science report*. Chestnut Hill, MA: Boston College.

Martin, M. O., Mullis, I., Gonzalez, E., O'Connor, Chrostowski, S., Gregory, K., Smith, T., & Garden, R. (2001). *Science benchmarking report TIMSS 1999 – Eighth grade.* Chestnut Hill, MA: Boston College.

McBride, J. W., & Silverman, F. L. (1991). Integrating elementary/middle school science and mathematics. *School Science and Mathematics, 91* (7), 285-292.

Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: A possible "Hidden Variable" in diagnostic pretest scores. *American Journal of Physics, 70* (12), 1259-1268.

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. A. (1998). *Mathematics achievement in Missouri and Oregon in an international context: 1997 TIMSS benchmarking*. Chestnut Hill, MA: TIMSS International Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., Smith, T. A. (2000). *TIMSS 1999: International mathematics report.* Chestnut Hill, MA: TIMSS International Study Center.

Mullis, I., Martin, M. O., Gonzalez, E., O'Connor, Chrostowski, S., Gregory, K., Garden, R., & Smith, T. (2001). *Mathematics benchmarking report TIMSS 1999 Eighth grade.* Chestnut Hill, MA: Boston College.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: The Author.

National Council of Teachers of Mathematics (2000). *Principles and standards of school mathematics.* Reston, VA: The Author.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Nixon, J. (1991). Reclaiming coherence: Cross-curriculum provision and the national curriculum. *Journal of Curriculum Studies, 23* (2), 187-192.

Orton, T., & Roper, T. (2000). Science and mathematics: A relationship in need of counseling? *Studies in Science Education, 35*, 123-154.

Ott, R. (1993). *An introduction to statistical methods and data analysis* (4th ed.). Belmont, CA: Duxbury.

Pang, J. & Good, R. (2000). A review of the integration of science and mathematics: Implications for further research. *School Science and Mathematics, 100* (2), 73-82.

Pashley, P. & Phillips, G. (1993). *Toward world-class standards: A research study linking international and national assessments.* Arlington, VA: National Science Foundation.

Underhill, R. G. (1995). Integrating math and science: We need dialogue! *School Science and Mathematics, 95* (2), 225.

Venville, G., Wallace, J., Rennie, L., & Malone, J. (1998). The integration of science, mathematics, and technology in a discipline-based culture. *School Science and Mathematics, 98* (6), 294-302.

Zabulionis, A. (2001, September 14). Mathematics and science achievement of various nations. *Education Policy Analysis Archives, 9* (33). Retrieved July 26, 2002 from http://epaa.asu.edu/epaa/v9n33/.