



University of Pennsylvania  
**ScholarlyCommons**

---

GSE Publications

Graduate School of Education

---

3-2003

# The Ecological Fallacy in Comparative and International Education Research: Discovering More From TIMSS Through Multilevel Modeling

Henry May  
*University of Pennsylvania*

Erling E. Boe  
*University of Pennsylvania, boe@gse.upenn.edu*

Robert Boruch  
*University of Pennsylvania, robertb@gse.upenn.edu*

Follow this and additional works at: [http://repository.upenn.edu/gse\\_pubs](http://repository.upenn.edu/gse_pubs)

 Part of the [Education Commons](#)

---

## Recommended Citation

May, H., Boe, E. E., & Boruch, R. (2003). The Ecological Fallacy in Comparative and International Education Research: Discovering More From TIMSS Through Multilevel Modeling. *CRESP Research Report*, Retrieved from [http://repository.upenn.edu/gse\\_pubs/410](http://repository.upenn.edu/gse_pubs/410)

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/gse\\_pubs/410](http://repository.upenn.edu/gse_pubs/410)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# The Ecological Fallacy in Comparative and International Education Research: Discovering More From TIMSS Through Multilevel Modeling

## **Abstract**

The ecological fallacy is the assumption that empirical relationships observed at the group level generalize to individuals within the groups, and vice versa, without empirical evidence supporting this assumption. When international data are analyzed, relationships can be uncovered not only at the student or school-levels, but also at the national-level. And the factors that explain differences between nations do not necessarily provide any information about the relationships between schools or students within those nations, or vice versa. Using data from TIMSS, several examples illustrating this point are presented, and the implications for comparative education research are discussed.

## **Disciplines**

Education

**THE ECOLOGICAL FALLACY IN COMPARATIVE AND INTERNATIONAL  
EDUCATION RESEARCH: DISCOVERING MORE FROM TIMSS  
THROUGH MULTILEVEL MODELING<sup>1</sup>**

**Henry May, Ph.D.**  
Research Specialist

**Erling E. Boe, Ph.D.**  
Professor of Education

and

**Robert F. Boruch, Ph.D.**  
University Trustee Professor of Education

Research Report No. 2003-TIMSS1

Center for Research and Evaluation in Social Policy  
Graduate School of Education  
University of Pennsylvania  
Philadelphia, PA 19104

March 26, 2003

---

<sup>1</sup>Support for this research was provided by a grant (Award Number R215U980021) from the Fund for Improvement in Education (CFDA Number: 84-215U), Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement, U.S. Department of Education; by a grant (Grant Number REC-9815112) from the Research on Education, Policy, and Practice Program (NSF 96-138), Division of Research, Evaluation, and Communication, Directorate for Education and Human Resources, National Science Foundation; and by the Center for Research and Evaluation in Social Policy, the Graduate School of Education of the University of Pennsylvania.



### **Abstract**

The ecological fallacy is the assumption that empirical relationships observed at the group level generalize to individuals within the groups, and vice versa, without empirical evidence supporting this assumption. When international data are analyzed, relationships can be uncovered not only at the student or school-levels, but also at the national-level. And the factors that explain differences between nations do not necessarily provide any information about the relationships between schools or students within those nations, or vice versa. Using data from TIMSS, several examples illustrating this point are presented, and the implications for comparative education research are discussed.



## TABLE OF CONTENTS

Abstract.....	iii
Table of Contents.....	v
Background.....	1
Methods.....	4
Data Sources.....	4
Dependent Variable.....	4
Independent Variables.....	5
Analysis Methods.....	7
Results.....	7
Discussion.....	11
References.....	15





## Background

Robinson (1950) first described the ecological fallacy when he demonstrated that the correlation between race and illiteracy becomes severely inflated when data are aggregated from the individual to the state level. He showed that the same phenomenon occurred in other analyses, and that the cause had mathematical rather than substantive origins. Yet at that time, ecological data, which represents aggregated characteristics of communities, was widely used in political science research and other social science arenas. Robinson's work "dismayed, and even infuriated many users of ecological data" (Alker, 1969).

Alker (1969) went further to describe a typology of inferential fallacies, showing that the ecological fallacy operates in more than one direction. He distinguished between fallacies of aggregation and fallacies of disaggregation. Over time, the term "ecological fallacy" came to describe any situation where a relationship described at one level is erroneously generalized to another. In education research, for example, this can happen when classrooms or schools are the units of analysis, and these aggregate results are used to explain differences between individual students, or when individual-level results are used to explain differences between classrooms or schools. Simply stated, relationships observed at one level may not validly be generalized to another level. To do so is regarded as a "fallacy." In fact, the relationships between the same variables might vary dramatically in size, and even change sign, when analyzed at individual and at group levels.

The ecological fallacy is widely recognized in the fields of political science and sociology. It is not so commonly acknowledged in education research, and the national context has only recently emerged as a potential unit of analysis in education research. The analyses outlined here show the importance of recognizing this methodological issue in comparative and international education research, with particular attention to national comparisons that have recently become possible in large-scale international education studies.

Since the publication of results and the public release of the data from the Third International Mathematics and Science Study 1995 (TIMSS), numerous other international education studies have been conducted or are still in process. The IEA has already released the data and results for the first repeat of TIMSS (Martin, Gregory, & Stemler, 2000), and is planning another replication in 2003 (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, et al., 2001) (see also

<http://www.timss.org> ). The IEA has also recently completed, or is currently conducting, four other large-scale studies including one focusing on civics (CIVED) (Lehmann, Schulz, Husfeldt, & Sibberns, 2001) and another on reading literacy (PIRLS) (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001) (see also <http://www.iea.nl> ). The OECD has also entered this arena with its own international study of reading, mathematics and science (PISA) (OECD, 2000) (see also <http://www.pisa.oecd.org> ).

All of these studies are large-scale, with most of them having 30 or more participating nations. The general purpose of these studies is to allow participating nations to gauge their standing relative to other countries' standings on achievement indicators, curriculum, student background and behaviors, and school and teacher characteristics. Termed "benchmarking," these comparisons involve only descriptive statistics at the national level. But this merely scratches the surface of what can be learned through analysis of international data sets. Through more rigorous exploration of relationships and trends that exist both within and across nations, the data from these studies can provide a much broader picture of factors that relate to educational inputs and outcomes worldwide.

The primary results of international achievement studies typically consist of two components: (a) the international comparison of mean national achievement scores and (b) the international comparisons of questionnaire data. The mean national achievement results usually receive the most attention and they might be metaphorically compared to those of a horse race. They show which nation "won" the race, along with the relative positions of each of the other nations. It was clear in the TIMSS 1995 results for eighth grade that the U.S. placed much further back in the pack than had been expected by most Americans. The question then was "why did U.S. students perform at such a mediocre level?"

The primary analysis of questionnaire data seemed to start to answer this question, showing mean achievement scores for students with particular questionnaire responses in each nation. But these results only implied relationships with achievement, and most of these were valid only for explaining differences in student performance within each nation, not national differences as evidenced by the horse race. Some secondary analyses of TIMSS data by independent researchers went further to provide more insight into the relationships that exist both within and across nations (Tarr, Mittag, Uekawa, & Lennex, 1999; Woessmann, 2001); however, important distinctions about the levels to which research results pertained often were not made. And if a

relationship that was observed within nations in primary or secondary analyses was used as an explanation for national-level differences or vice versa, the “ecological fallacy” might result. In fact, analyses of the TIMSS 1995 data presented herein suggest that relationships that differ substantially from the student to the national level are not uncommon.

One way to avoid being deceived by the ecological fallacy is to analyze and interpret variables only at the levels at which they are measured. Unfortunately, this results in the loss of important information about the characteristics of groups. Aggregation of individual-level variables in order to describe groups is often helpful. In international studies, student-level data are often aggregated to the national level in order to characterize the students in a nation as a group. For example, TIMSS students were asked to provide information about their parents’ education. At the student level, there is a clear positive relationship between parents’ education and student achievement at the individual level. Additionally, the typical level of education of parents in each nation (e.g. the national average) is easily calculated, and this characterization of national educational attainment is positively related to national average achievement (May, 2001). The most informative research on the relationship between student achievement and parents’ educational attainment would involve analysis at both the student and the national level, possibly comparing the strength of the relationship at the two levels. Furthermore, the availability of new software that can perform this type of analysis makes it a feasible methodology for most researchers.

It is important to note that, in some cases, national aggregation of a student-level variable can produce a different variable at the national level which, unlike the original variable, conveys information about the national context or the “environment” in that nation. For example, a student’s level of agreement with the statement “Math is easy” shows their perception of the level of ease with which they learn mathematics relative to their peers. In contrast, a national aggregation of students’ responses to the same question shows the typical perceived level of ease with which mathematics is learned in that nation. The student-level variable might well represent individual ability relative to other students in the same nation, whereas the national-level variable is more likely to reflect the rigor of a nation’s mathematics curriculum. Clearly, individual ability and the rigor of the curriculum have the potential to drive both indicators, and the following four assumptions are necessary to support such an interpretation:

1. the ability to learn mathematics varies from student to student
2. the rigor of mathematics curriculum varies from nation to nation
3. students' ability to learn mathematics varies little from nation to nation
4. the rigor of mathematics curriculum varies little within each nation

At least the first three of these assumptions are plausible. A violation of the fourth assumption would alter the meaning of the student-level variable to also reflect the rigor of the local curriculum. Nevertheless, it is likely this student-level variable and this national-level variable, although derived from the same survey item, represent very different constructs. It is under these circumstances that understanding the ecological fallacy becomes very important.

The purpose of this paper is to illustrate, using actual TIMSS data, how the ecological fallacy might occur when conducting international or comparative analyses. The relationships between mathematics achievement and seven variables measured at the student level are compared at the individual-level and when aggregated to the national level. Each comparison illustrates (a) how relationships can differ between levels in size, direction, or both size and direction, and (b) how the interpretations of variables describing national contexts can be very different from the interpretations of similar student-level variables.

## Methods

### *Data Sources*

These analyses use TIMSS 1995 student background questionnaire data and mathematics achievement scores for over 147,000 eighth grade students in 41 nations, weighted to represent national populations of students. See Martin and Kelly (1996, 1997) for a technical description of TIMSS.

### *Dependent Variable*

The dependent variable in each model is the 8<sup>th</sup> grade TIMSS mathematics achievement score at the student level. These are internationally scaled IRT scores, with an international mean of 500 and standard deviation of 100. See Martin and Kelly (1996, chapters 6 & 7) for a full description of the TIMSS tests and scaling methodology.

### *Independent Variables*

In order to demonstrate the range of the ecological fallacy threat, seven variables were selected based on the size and direction of their within-nation (student level) and between-nation (national level) relationships. The independent variables were derived from students' responses to items on the TIMSS student background questionnaire. These variables are described in detail in Table 1. Survey responses for three of the variables were on a four-point Likert scale of agreement. The distribution of responses for all three variables was heavily skewed toward "agree" and "strongly agree." In fact, in most of the nations in TIMSS, over 70% of students selected "agree" or "strongly agree" in response to these three items. To improve both the symmetry of the distribution of responses and the interpretability of the regression estimates, each of these three variables were dichotomized near the middle of their distributions<sup>1</sup>. This produced dummy variables at the student level comparing students responding "strongly agree" to all others ("strongly agree"=1; "agree," "disagree," and "strongly disagree"=0) and proportions of "strongly agree" responses at the national level. Three more variables were measured on a five-point time scale. These variables were also rescaled, using interpolation, to improve the interpretability of the regression estimates ("no time"=0, "less than 1 hour"=.5, "1-2 hours"=1.5, "3-5 hours"=4, "more than 5 hours"=7). Responses on the one remaining variable were yes/no ("yes"=1, "no"=0).

After rescaling the items, means for each independent variable were calculated for each nation. These national means were then subtracted from their respective student-level scores, producing group-mean centered scores at the student level. Therefore, each student's score represented his/her response position on an independent variable relative to the responses of other students in his/her nation. The variation in student responses on the independent variables that was attributable to nation-level differences was removed from the student-level predictors. This nation-specific information was now represented solely by the national-level means.

---

<sup>1</sup> Dichotomization of these variables using an alternate coding ("agree/strongly agree"=1; "disagree/strongly disagree"=0) results in a very skewed binary variable, comparing the vast majority of students (those responding "agree" or "strongly agree") to a small group of students (less than 10% responded "disagree" or "strongly disagree" in some nations).

Table 1. Student Questionnaire Items Used to Derive Independent Variables with Variable Codings.

Survey Question	Survey Item	Response Scale and Student-Level Coding	National-Level Coding
I enjoy learning mathematics.	23A	Strongly Agree = 1 Agree = 0 Disagree = 0 Strongly Disagree = 0	Proportion of students who marked "strongly agree"
I usually do well in mathematics.	17A	Strongly Agree = 1 Agree = 0 Disagree = 0 Strongly Disagree = 0	Proportion of students who marked "strongly agree"
My mother thinks it is important for me to be good at sports.	13D	Strongly Agree = 1 Agree = 0 Disagree = 0 Strongly Disagree = 0	Proportion of students who marked "strongly agree"
On a normal school day, how much time do you spend before or after school reading a book for enjoyment?	6F	No time = 0 Less than 1 hour = .5 1-2 hours = 1.5 3-5 hours = 4 More than 5 hours = 7	Average number of hours per day
During the week, how much time before or after school do you usually spend at a paid job?	5D	No time = 0 Less than 1 hour = .5 1-2 hours = 1.5 3-5 hours = 4 More than 5 hours = 7	Average number of hours per day
On a normal school day, how much time do you spend before or after school watching television and videos?	6A	No time = 0 Less than 1 hour = .5 1-2 hours = 1.5 3-5 hours = 4 More than 5 hours = 7	Average number of hours per day
Do you have a computer at your home?	12B	Yes = 1 No = 0	Proportion of students who marked "yes"

Note. Items from the Population 2 Student Questionnaire administered as part of the Third International Mathematics and Science Study (TIMSS-1995).

### *Analyses Methods*

Hierarchical linear modeling (HLM) was used to explore, simultaneously, within and between-nation relationships of students' math achievement scores and the independent variables derived from the student background questionnaire. Each HLM model contains two levels: student and nation<sup>2</sup>. The seven independent variables were examined separately. This produced (a) estimates of the variance partition of student achievement between the student and national levels, and (b) estimates of the size and direction of relationships between each predictor variable and achievement at both the student (N=147,505) and the national level (N=41). The sample sizes for each analysis were reduced slightly due to missing data resulting from item non-response and national variations in the student background questionnaires. The analysis with the smallest available sample was based on 133,338 students in 40 nations.

### **Results**

The unconditional HLM model (without any predictors) showed that a significant ( $p < .0001$ ) and substantial portion of variability in student mathematics achievement at grade 8 is attributable to the national-level. Specifically, 31% of the variance in mathematics achievement exists at the national level; the remaining 69% exists within nations.

Table 2 shows the results of the HLM models for each of the seven independent variables. In order to ensure comparability of estimates, the size of each relationship is represented as the percentage of the variability in mathematics achievement explained by each predictor at the student and national levels. Across the seven variables, the sizes of relationships are consistently much smaller at the student level than at the national level. The percentages of student-level variance explained ranged from 0.9% to 6.0%. The percentages of national-level variance explained ranged from 9.5% to 56.1%.

The seven predictor variables selected for analysis show wide differences in their relationships with math achievement at the student level and the national level. Therefore, generalizing results from one level to the other for any of these seven variables would constitute the ecological fallacy. Three types of results were found that preclude valid cross-level generalizations: (a) a positive

---

<sup>2</sup> A contextual model, using within-group centering was used to estimate, simultaneously, the effects at the two levels. See Bryk & Raudenbush, 1992; Cronbach, 1977; and Iversen, 1991 for examples of and discussion of contextual models.

**Table 2. Size and Direction of Relationships for Seven Independent Variables Predicting Eighth Grade Mathematics Achievement.**

Independent Variable	Sample Size		Percent Variance Explained and Direction of Effect (+,-)	
	Student	Nation	Student	Nation
enjoy learning mathematics	140,204	40	0.9(+)	56.1(-)
usually do well in mathematics.	143,732	41	6.0(+)	31.3(-)
mother thinks sports are important	135,465	40	1.1(-)	27.1(-)
time spent reading a book for enjoyment	140,650	41	0.4(+)	17.3(-)
time spent at a paid job	133,338	40	1.1(-)	26.1(-)
time spent watching television and videos	142,354	41	1.3(-)	9.5(+)
has a computer at home	138,996	40	1.9(+)	12.6(+)

linear relationship between a predictor variable and math achievement at the national level, with a negative linear relationship at the student level, (b) a negative linear relationship at the national level, with a positive linear relationship at the student level, and (c) a large positive or negative linear relationship at the national level, with a small relationship in the same direction at the student level.

A visual depiction of the results is shown in Figure 1, with each of the seven independent variables plotted in one of four quadrants. The axes in this plot show simultaneously the strength and direction of relationships within nations (horizontal) and across nations (vertical). Quadrants I and III show variables that have the same direction of relationships at the student and national level. Quadrants II and IV show variables that have the opposing directions of relationships. Note that all seven of the variables plotted here show a larger relationship at the national level than at the student level, regardless of quadrant.

Figure 2 shows a positive linear relationship between students' responses to "I usually do well in math" and their math achievement scores within the United States ( $r = .29$ ), and a negative



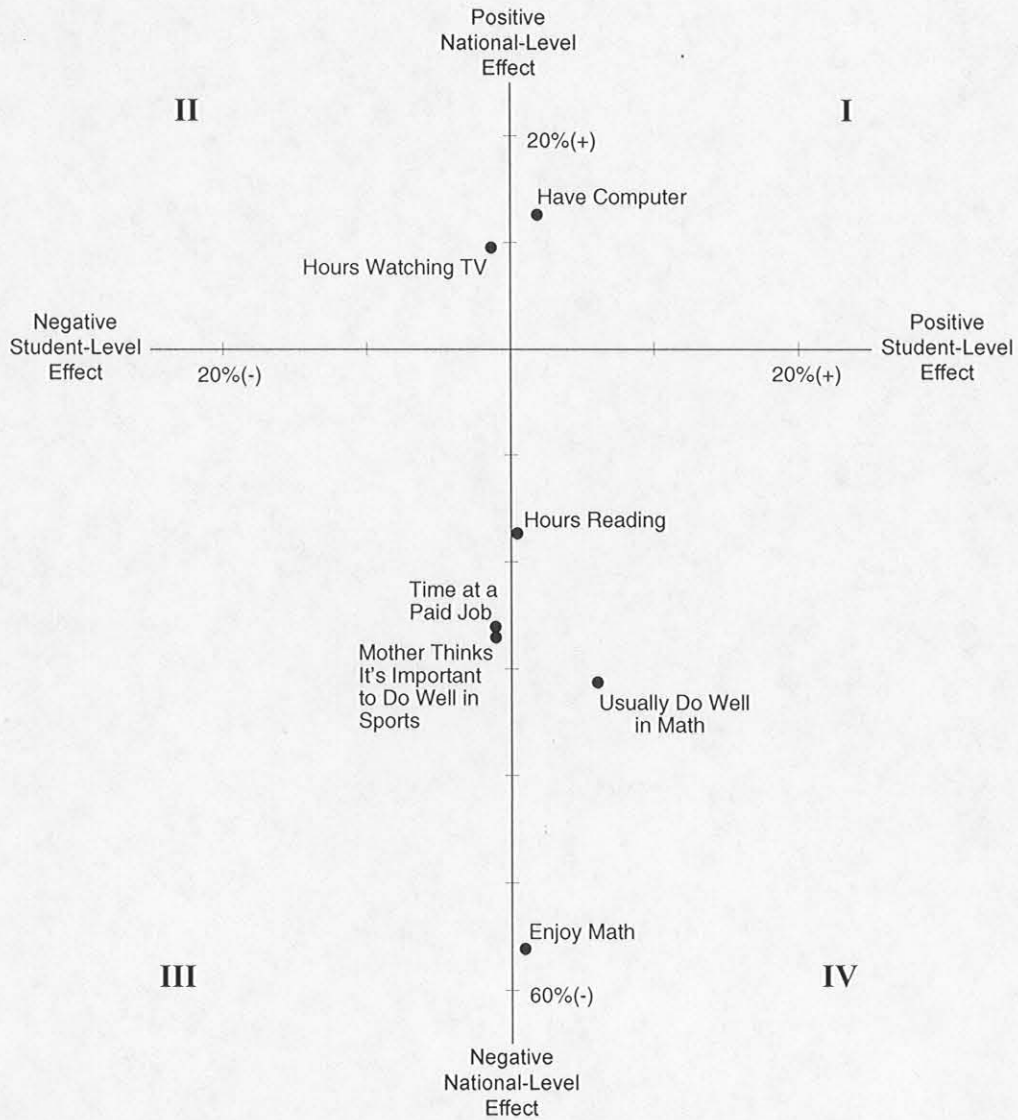


Figure 1. National and student-level effect sizes (percent variance explained) and effect direction for seven contextual variables.

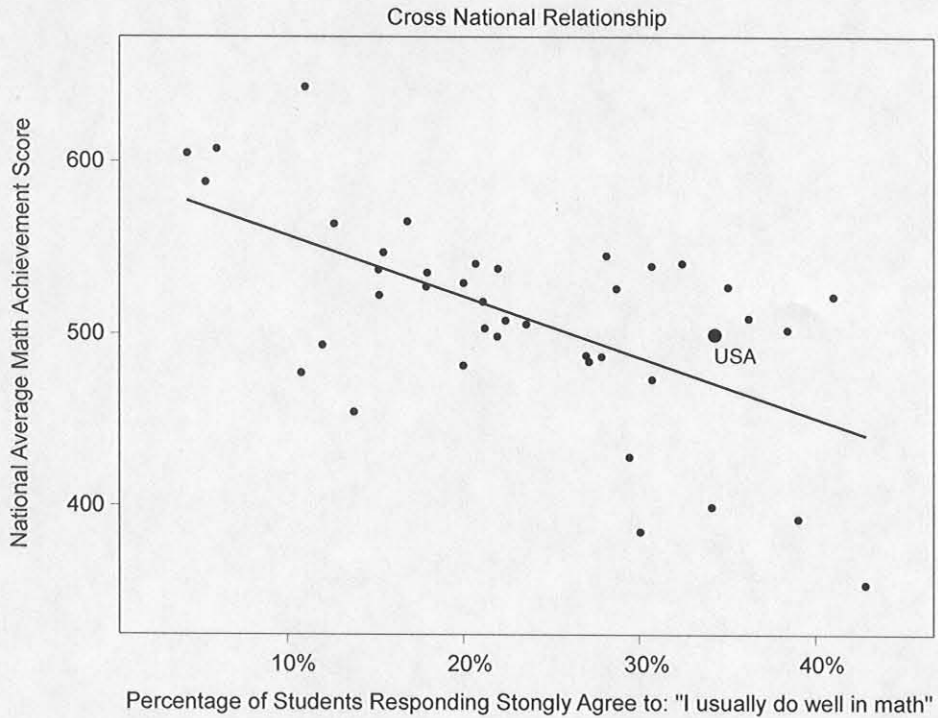
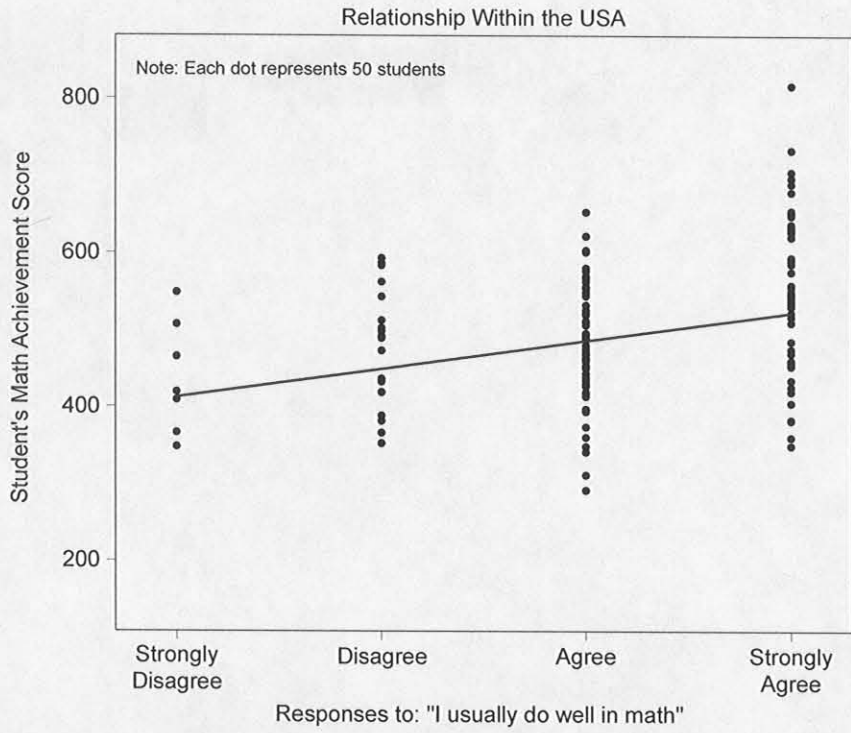


Figure 2. Within versus between-nation effects of the "usually do well" variable.

relationship at the national level ( $r = -.56$ ). The United States has the seventh highest proportion of students responding “strongly agree.”

Figure 3 shows a negative linear relationship between “hours worked at a paid job” within Switzerland and also cross-nationally. However, the national-level relationship is much stronger ( $r = .51$ ) than the relationship within Switzerland ( $r = -.10$ ).

### Discussion

The simple message of the multilevel analyses presented here is that comparative educators should be acutely aware of the potential for committing the ecological fallacy, i.e., to assume that research findings at the cross-national level also represent within-nation relationships, and vice versa. These results also demonstrate the value of the national level in education research, and that many factors that influence student achievement warrant completely different interpretations when analyzed at different levels. If we assume that contextual effects do not exist, we risk committing the ecological fallacy. Researchers and consumers of research are particularly vulnerable to this fallacy if an analysis is performed only at one level, with the temptation to generalize to another level. Therefore, it is very important to identify groups and nesting in educational data, and to perform analyses that recognize levels of aggregation where contextual differences occur.

While these results clearly demonstrate that numerical differences in relationships are likely to occur when a variable is analyzed at multiple levels, they also show the importance of recognizing that the interpretation of relationships for the same variable might be dramatically different in different contexts. For example, watching more TV shows a detrimental effect on an individual student’s achievement relative to other students in his or her nation. However, the cross-national effect of more TV is positive. This national-level relationship is likely to be confounded with national wealth; however, if such bivariate national-level results were not as easily dismissed as spurious, and were presented without the student-level effects, they might guide education policy in a very wrong direction.

Analogously, in the case of the “usually do well in math” variable, the interpretations at these two levels are vastly different. A positive relationship is evident within nations, suggesting that students are reasonably able to gauge their level of achievement relative to the performance of their peers. Simply stated, within each nation, students who say they do well in math generally

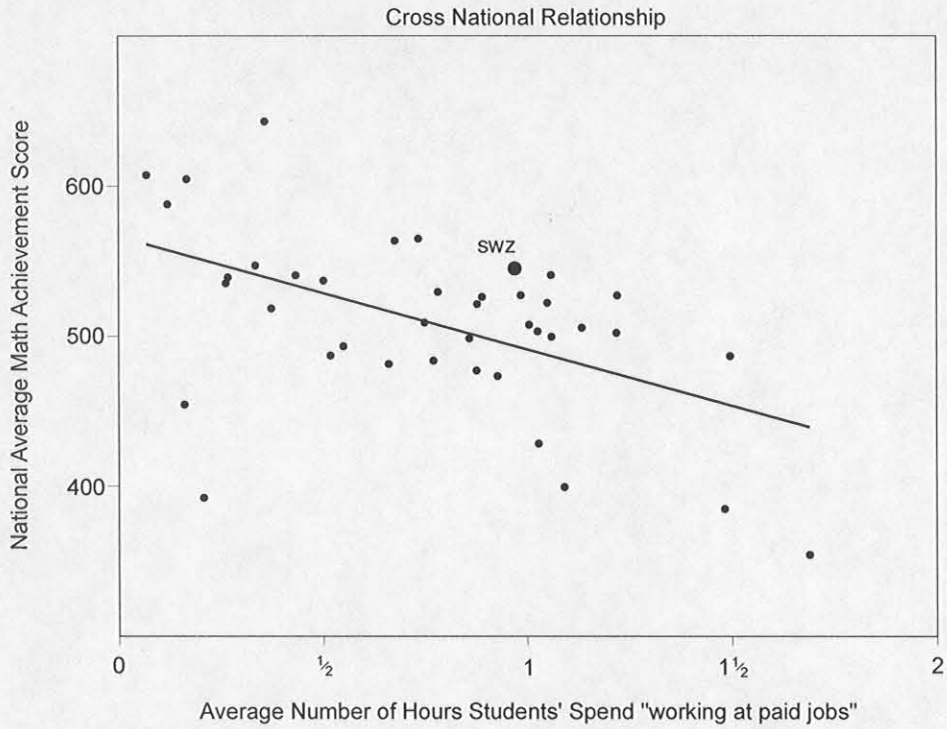
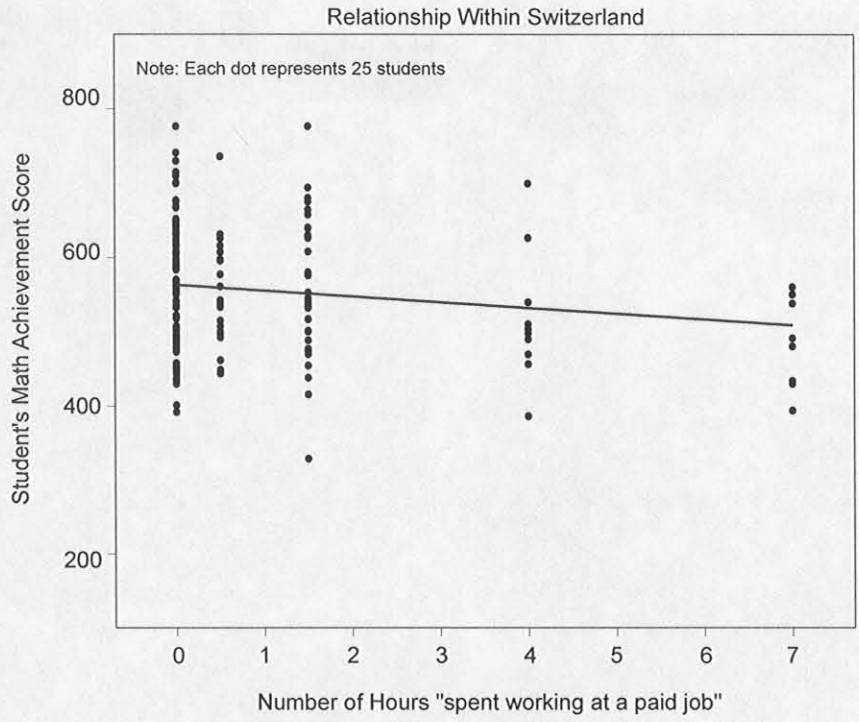


Figure 3. Within versus between-nation effects of the "time spent at a paid job" variable.

have higher achievement scores than students who say they do not do well. By contrast, a negative relationship is observed across nations, suggesting that, on average, students in poor performing nations may be likely to overestimate their own achievement relative to an international standard. This can be attributed to a lack of exposure to international standards and the location of the context-specific baseline within each nation to which students compare themselves.

Much like the example involving students' responses to "Math is easy," the responses to "I usually do well in Math" are likely to be driven by both individual student ability and the rigor of the curriculum. However, the "do well" item may also be driven by the performance expectations of students, teachers, and parents in a nation. If a grade of "D" is acceptable, then "D" students might report that they usually do well in mathematics. Therefore, it is possible that performance expectations are lower in poor performing countries and higher in high performing countries. It may also be the case that in poor performing countries, most students perform at a mediocre level, lowering educational standards. Or it may be the case that educational standards in poor performing countries are themselves low, and the vast majority of students find it very easy to meet the standard – every student is an "A" student.

Students' responses to items like "I usually do well in mathematics" potentially contain much more information than what is initially apparent. When variables are aggregated to the national level, they can take on new meanings and convey information about the national context that had previously been hidden. Such aggregated variables often convey information about the culture of a nation, for they represent the general condition of the people in a nation. This kind of knowledge is invaluable in comparative and international education research (as should be the distinction between relationships attributable to the national level and those observed within-nations).

Educational researchers around the world have analyzed differences between individual students, classrooms, schools, and other levels of educational systems. As demonstrated by these analyses, the importance of the national context is clear. By including nation and other important levels as units of analysis in comparative and international studies, we increase the validity of our inferences and reduce the potential for being fooled by the ecological fallacy.



## References

- Alker, H. R. (1969). A Typology of Ecological Fallacies. In M. Dogan and S. Rokkan eds. Quantitative Ecological Analysis in the Social Sciences. Cambridge, MA: MIT Press.
- Bryk, A. S. and Raudenbush S. W. (1992). Hierarchical Linear Models. Newbury Park, CA: Sage.
- Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). Framework and Specifications for PIRLS 2001 Assessment. Chestnut Hill, MA: International Study Center, Boston College.
- Cronbach, L. J. and Webb, N. (1975). Between-Class and Within-Class Effects in a Reported Aptitude x Treatment Interaction: Reanalysis of a study by G. L. Anderson. Journal of Educational Psychology, 67(6), pp. 717-724.
- Iversen, G. R. (1991). Contextual Analysis. Newbury Park, CA: Sage.
- Lehmann, R., Schulz, W., Husfeldt, V. and Sibberns, H. (2001): Technical report: IEA Civic Education Study. Amsterdam: IEA.
- Martin, M. O., & Kelly, D. L. (Eds.) (1996). Third International Mathematics and Science Study Technical Report: Volume I: Design and Development. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Martin, M. O., & Kelly, D. L. (Eds.) (1997). Third International Mathematics and Science Study Technical Report: Volume II: Implementation and Analysis (Primary and Middle School Years). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Martin, M. O., Gregory, K. D., & Stemler, S. E. (Eds.) (2000). TIMSS 1999 Technical Report. Chestnut Hill, MA: International Study Center, Boston College.
- May, H. (2001). Socioeconomic Status and TIMSS: Derivation and Analysis of a Composite Scale. Unpublished manuscript.
- Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., Chrostowski, S. J., & O'Connor, K. M. (2001). TIMSS Assessment Frameworks and Specifications 2003. Chestnut Hill, MA: International Study Center, Boston College.

OECD, Organization for Economic Co-Operation and Development. (2000). Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy. Paris: OECD.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. American Sociological Review, 15, pp. 351-357.

Tarr, J. E., Mittag, K. C., Uekawa, K., & Lennex, L. (1999, April). A comparison of calculator use in eighth-grade mathematics classrooms in the United States, Japan, and Portugal: results from the third international mathematics and science study. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Woessmann, L. (2001). Why students in some countries do better: international evidence on the importance of education policy. Education Matters, 1(2), pp.67-74.