The New Educational Review

Krzysztof Konarzewski
Poland

Primary School Pupils' Attitude toward Mathematics and Their Achievement in Mathematics

DOI: 10.15804/tner.2019.56.2.01

Abstract

International data from TIMSS 2015 show a significant negative correlation between the country mean attitude toward mathematics and the mean achievement in mathematics of fourth grade pupils. The aim of the paper is to decide whether it is a statistical artifact or an indication of a real factor operating at the country level. The multilevel regression analysis of the data attests the latter. The factor is hypothetically identified with country typical pressure for knowledge acquisition. Strong pressure is conducive to high achievement but it puts pupils under stress, which lowers the attitude. The reverse holds for weak pressure. Within country, the variability of pressure for knowledge acquisition is restricted, hence pupils may maintain psychological coherence between achievement and attitude.

Keywords: *primary education, achievement in mathematics, attitude toward mathematics, TIMSS*

Introduction

Typically, a research paper begins with a sketch of a theory and the hypotheses which logically follow from it. This paper is different as it begins with an intriguing data pattern shown in Figure 1.

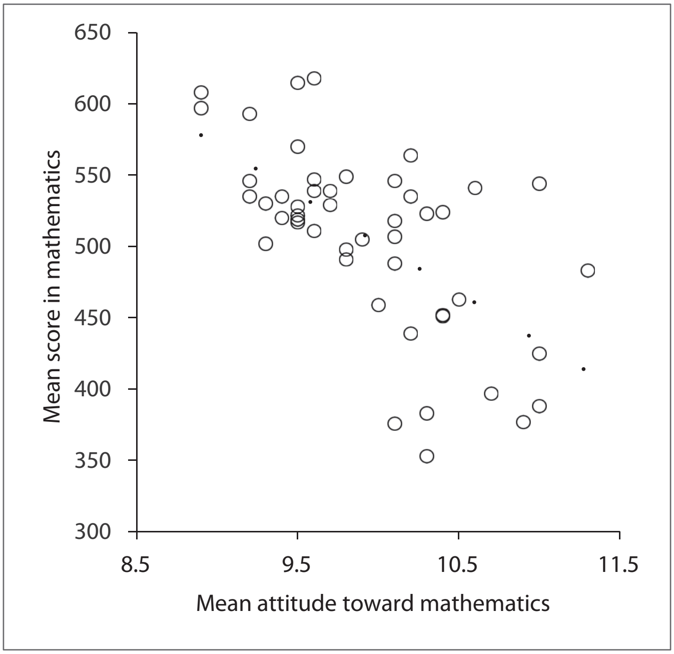

Figure 1. The regression of the country mean score in mathematics on the country mean attitude toward mathematics. Source: Mullis et al., 2016

Data were taken from TIMSS 2015 – the latest edition of the Trends in International Mathematics and Science Study. TIMSS is a program of cyclical measurements of scholastic achievement in mathematics and science of ten year old pupils. In 2015, 49 countries from throughout the world participated in the study. The international report (Mullis et al., 2016) provides, among other things, two values for each country: the mean score in the mathematics test and mean score in the scale of attitude toward mathematics. Each country may be represented by a point in the Cartesian axes. Figure 1 shows that there is a strong correlation ($r = -0.62$) between two aggregates: the more favorable the mean attitude toward mathematics, the lower the mean score of mathematics. How can this correlation be understood? Could it be that mathematics becomes increasingly more disappointing for a pupil as he/she delves into it deeper? Whoever reasons in this way is making an ecological fallacy.

Ecological fallacy

An ecological fallacy emerges whenever we infer something about individuals from the data collected for the groups to which these individuals belong. An instance of this fallacy is provided in a seminal paper by William Robinson (1950). For each state of the United States, he computed the proportion of illiterate persons aged 10 and more, and the proportion of people born outside the US. He found that the correlation of the two aggregates was $r = -0.53$. The higher the percentage of immigrants, the lower the percentage of illiterate persons. May we infer from this that the probability of being illiterate was higher for natives than for immigrants? No, since the total correlation (r) of individual data was not only lower, but also went in the opposite direction (0.12).

The weighted correlation between the m pairs of X- and Y-aggregates (means, medians, or percentages) which describe the groups was called by Robinson an ecological correlation (re). He was able to show that re is equal to r only if the weighted average of the m within-group individual correlations between X and Y (rw) is not less than r, which is never the case. He also showed that re is numerically greater than r whenever rw is not greater than r, which is usually the case. Finally he explained why the numerical value of re increases as m decreases, i.e., as we consolidate many smaller groups into a few larger ones. He concluded that ecological correlations cannot be used validly as substitutes for individual correlations and recommended the use of meaningful correlations between the properties of individuals, instead of meaningless ecological correlations when studying similar problems. Both the conclusion and the recommendation met with serious criticism.

One line of criticism is well represented by Gove and Hughes (1980). They reported that living alone was a powerful predictor of suicide and alcoholism in US cities of over 50 thousand inhabitants. The method they used consisted of regressing suicide and alcoholism rates on the proportion of people living alone in each of the cities. The main threat to the validity of the conclusion poses the possibility that there are some personality traits (e.g., depression) which may push people both to living alone and to suicide. However, if the size and density of the cities as well as crucial social and economic aggregates are directly controlled for, then the city proportion of persons with these traits may be considered random. For some problems, aggregated data may then be superior to individual data, providing that the model is properly specified and all relevant confounding variables are properly controlled for.

Another line of criticism was initiated by Hayward Alker (1969). He supplemented ecological fallacy with "individualistic fallacy", which is committed whenever "ideologically motivated social scientists try to generalize from individual behavior to collective relationships" (p. 78). Subramanian et al. (2009) argue that to avoid both fallacies, a multilevel approach must be taken. This allows for an examination to be made of the individual outcome circumstances being shaped by higher-level factors, including the compositional effects of individual variables. To demonstrate the potential fallacies of considering relationships at only one level, the authors used the Robinson data and modeled the log odds of being illiterate in three groups (native whites, foreign-born whites and blacks) nested within 49 states. Similarly as in Robinson's paper, foreign-born whites (and blacks) have a 5–6 times greater chance to be illiterate than native whites, but the individual relationship between race/nativity and illiteracy varied across states. The biggest state variation was observed for native whites and the smallest for foreign-born whites. Contrary to Robinson's paper, a state's race/nativity composition did not account for the state variation in illiteracy. However, the authors found another state-level factor – segregation or desegregation in the education system (Jim Crow laws) – which strongly influenced the individual odds of being illiterate, controlling for individual-level race/nativity. Especially blacks in segregated-school states were almost 30 times more likely to be illiterate than native whites residing in desegregated-school states. With the approach recommended by Robinson, the effects of Jim Crow laws would never be detected.

Research Methodology

The aim of the paper is to decide whether the effect shown in Figure 1 is a statistical artifact or rather an indication of a real factor operating on the country level. Following Subramanian's advice, it will be done by means of multilevel regression analysis on the TIMSS 2015 data.

Measurements

1. Scale of attitude toward mathematics. Pupils responded to 28 items in a four-point scale: Agree a lot – Agree a little – Disagree a little – Disagree a lot. For 21 items, an "Agree a lot" response was scored at four points and a "Disagree a lot"

response – one point. Seven items were scored in reverse order. Out of these items, IEA experts constructed three scales of attitude:

- toward mathematics as a subject of study (e.g., Mathematics is one of my favorite subjects),
- toward a teacher of mathematics and his/her lessons (e.g., My teacher is good at explaining mathematics),
- toward oneself as a learner of mathematics (e.g., I am good at working out difficult mathematics problems).

Unfortunately, the three scales are not independent (Table 1) and this poses a collinearity problem.

Table 1. Pearson's correlation coefficients

Attitude	Toward a subject	Toward a teacher	Toward oneself
Toward a subject		0.47 (0.002)	0.49 (0.002)
Toward a teacher	0.67 (0.01)		0.23 (0.002)
Toward oneself as a learner	0.58 (0.01)	0.48 (0.02)	

Note: Above the diagonal, there are the weighted coefficients from international data (n = 247,344) and in parentheses standard errors estimated by the bootstrapping method. Below the diagonal, there are the weighted coefficients from the Polish data (n = 4,723) and in parentheses standard errors estimated by the jackknife method.

Instead of splitting hairs, it is better to settle on one scale. In fact, Cronbach's internal consistency statistics for all 28 items is 0.93 and the first principal component explains 36% of variance. The set of items seems unidimensional and may be scaled by means of the IRT. Before doing so, however, let us check the distribution of raw scores. The distribution is peculiar: its mode is identical with the maximum value. As many as 9,931 "enthusiastic" pupils (4% of the total) responded to all of the items in a way that yielded four points. On the other hand, there were only ten pupils who exclusively used the one-point options. It should be remembered that since the coding of seven items was reversed, the mechanical selection of the left- or right-side options could not account for the mean score of one or four. A pupil with such scores would have to read and understand each item. We have to understand this peculiarity before going further.

Had the enthusiasts been randomly distributed among the classes, they could have been found to be psychologically peculiar. But the distribution is not random (Table 2). When there is more than 10% of enthusiasts in a class, we may reasonably suspect that their enthusiasm was manipulated, i.e., that the children were

Table 2. Distribution of enthusiasts' share in classes

Percent of enthusiasts in a class	Number of classes	Percent of classes
0	7,354	56.7
0.1–10	3,917	30.2
10.1–20	1,346	10.4
20.1–30	250	1.9
30.1–40	69	0.5
40.1–50	11	0.1
50.1–60	12	0.1
60.1–70	2	0.0
80.1–90	2	0.0
Over 90	1	0.0
Total	12,964	100.0

pressured by adults (parents, teachers or pollsters) to manifest their satisfaction with learning and teaching. In some countries, there are more such classes than in others, e.g., 40% in Bulgaria and none in Korea (Table 3). If the mean pressure is negatively correlated with the mean achievement, it might explain the effect shown in Figure 1. The antecedent is true, but the consequence is false. After removing all the enthusiasts, the aggregates are still negatively correlated ($r = -0.64$).

Table 3. Percent of classes by country in which pressure might have been exerted on pupils, and the mean score of the mathematics test

Country	Per-cent	Mean test score	Country	Percent	Mean test score
Bulgaria	40.0	524	France	9.8	488
Turkey	33.6	483	Norway	9.6	549
Cyprus	32.4	523	Germany	9.4	522
Serbia	32.4	518	Lithuania	9.1	535
Iran	31.9	431	Croatia	8.9	502
Kazakhstan	28.3	544	Slovak Republic	8.4	498
Qatar	27.5	439	Australia	8.0	517
Bahrain	27.1	451	Russian Federation	7.9	564
Georgia	19.9	463	Northern Ireland	7.8	570

Country	Per-cent	Mean test score	Country	Percent	Mean test score
Kuwait	19.6	353	Poland	5.5	535
Saudi Arabia	19.6	383	Ireland	5.4	547
Morocco	19.3	377	Chile	5.1	459
Oman	19.1	425	Denmark	4.7	539
Hungary	18.0	529	New Zealand	4.7	491
Spain	15.0	505	Netherlands	3.6	530
Indonesia	14.0	397	Singapore	3.6	618
United Arab Emirates	13.5	452	Czech Republic	2.8	528
Portugal	13.4	541	Hong Kong	2.8	615
United States	12.9	539	Sweden	2.4	519
Slovenia	12.4	520	Belgium (Flemish)	2.1	546
Italy	10.4	507	Finland	1.5	535
England	10.4	546	Chinese Taipei	1.1	597
Canada	10.3	511	Japan	0.0	593
			Republic of Korea	0.0	608

The attitude toward mathematics was eventually scaled using a graded response IRT model. The distribution of the standardized variable is slightly censored from the right with a skew –0.20 and kurtosis –0.18. It correlates with the original scales: 0.93 with the scale toward the subject and 0.65 toward both the teacher and oneself.

2. Test of mathematics. The instrument consists of 169 items distributed among 14 booklets. The dichotomous items were scaled in two- and three-parameter IRT models, while the polychotomous items – in the partial credit model. The scale was aligned with the scale of TIMSS 1995 with a mean of 500 and a standard deviation of 100. Five plausible variables were computed for each pupil. They jointly entered the subsequent analyses, allowing the inclusion of a standard error of measurement to be made into the standard error estimates.

3. Home resources scale. This is an IRT standardized composite of four variables: the number of children's books at home, the child has their own room with Internet connection, the level of parental education, and the level of parental occupation. The scale is used here as an approximation of the socioeconomic status of the pupil's family.

Table 4. Descriptive statistics

Variable	*N*	Min	Max	Mean	St. deviation
Achievement in mathematics	237,403	97.4	837.7	505.5	96.6
Attitude toward mathematics	237,403	–4.38	2.05	0.00	1.00
Home resources	190,951	–3.47	2.59	0.00	1.00
Pupil's age (years)	237,077	6.37	14.98	10.18	0.59

Analysis

Analysis was carried out by means of a three-level hierarchical linear model (Raudenbush and Bryk, 2002), defined as follows.

$$(\text{Mathematics score})_{jk} = \pi_{0jk} + \pi_{1jk}\,(\text{Attitude})_{ijk} + \pi_{2jk}\,(\text{Gender})_{ijk} + \pi_{3jk}\,(\text{Age})_{ijk} + \pi_{4jk}\,(\text{Home resources})_{ijk} + e_{ijk} \qquad \text{[Level 1]}$$

In this equation, $\pi_0$jk represents the achievement mean of class *j* from a country *k*. π_{1jk} tells us to what degree the mathematics scores are dependent on the pupil's attitude toward mathematics. Gender, age and home resources are covariates.

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\,(\text{Class mean attitude})_{jk} + \beta_{02k}\,(\text{Class mean age})_{jk} + \beta_{03k}\,(\text{Class mean of home resources})_{jk} + r_{0jk} \qquad \text{[Level 2]}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k}$$

$$\pi_{3jk} = \beta_{30k}$$

$$\pi_{4jk} = \beta_{40k} + r_{4jk}$$

In the first equation, β_{00}k represents the achievement mean in country *k*. β_{01}k tells us to what degree the mean achievement of a class is dependent on the mean pupils' attitude toward mathematics. Mean of age and home resources of a class are covariates, with the latter accounting for the social milieu of the school. No distinction is made between class and school level because a considerable number of schools (31% in Poland) have only one fourth grade class.

$\beta_{00k} = \gamma_{000} + \gamma_{001}$ (Country mean attitude)$_k$ + γ_{002} (Country mean age)$_k$ + γ_{003} (Country mean of home resources)$_k$ + u_{00k} [Level 3]

$\beta_{01k} = \gamma_{010} + u_{01k}$

$\beta_{02k} = \gamma_{020} + u_{02k}$

$\beta_{03k} = \gamma_{030} + u_{03k}$

$\beta_{10k} = \gamma_{100} + u_{10k}$

$\beta_{20k} = \gamma_{200} + u_{20k}$

$\beta_{30k} = \gamma_{300} + u_{30k}$

$\beta_{30k} = \gamma_{400} + u_{40k}$

In the first equation, γ_{000} represents the international achievement mean, γ_{001} is a measure of the dependency of mean achievement in a country on the country weighted mean attitude. Weighted mean of age and home resources of a country are covariates.

Subjects

The international data base has records of 253,371 pupils from 13,462 classes in 47 countries (Data from Jordan and South Africa are unavailable). Two countries did not submit data on home resources. Further deletions consisted of 2,510 pupils who did not complete more than half of the attitude items, 9,931 "enthusiasts", 539 classes with fewer than five pupils, and 27 classes in which the attitude scale was not administered. After all of the reductions, data of 190,690 pupils (49.4% of girls) from 12,128 classes in 45 countries remained.

Results

Within classes, a commonsense tenet holds: regardless of gender, age, and home resources for learning, achievement and attitude go hand in hand (Table 5). A correlational study cannot decide which of the two is a cause and which is an effect. Besides, this may be a chicken and egg situation. An interest in mathematics may stimulate pupils to learn mathematics and their success in learning may further strengthen their interest. Please note that in an average class, there is no birth date effect, but there are effects of gender (to the advantage of boys) and home resources.

Table 5. Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	se	*p*
International mean achievement, γ_{000}	509.42	9.71	<0.001
International attitude slope, γ_{001}	–91.65	30.61	0.005
International age slope, γ_{002}	38.10	15.68	0.020
International home resources slope, γ_{003}	27.62	27.84	ns.
Class attitude slope, γ_{010}	15.29	2.33	<0.001
Class age slope, γ_{020}	12.83	4.97	0.004
Class home resources slope, γ_{030}	52.88	2.99	<0.001
Individual attitude slope, γ_{100}	16.26	0.62	<0.001
Individual gender slope, γ_{200}	2.47	1.09	0.029
Individual age slope, γ_{300}	0.70	1.70	ns.
Home resources slope, γ_{400}	23.47	1.19	<0.001

Note: Unweighted data.

At the between-class level, we basically note a similar pattern: independently of the average age and home resources, pupils have more favorable attitudes towards mathematics in classes with higher mean achievement. At the international level, however, the reverse holds true: in countries with higher mean achievement, pupils have less favorable attitudes. A drop-off in the country mean attitude by one standard deviation is associated with an increase in achievement by 92 points. The attitude variable reduces the variance of individual residuals ($e$ijk) by 5.3% to a value of 3774.5. The variances of class ($r_0$jk) and country (u_{00}k) residuals are reduced by 4.3% and 14.9%, respectively.

Conclusion

The negative correlation between countries' average achievement in mathematics and average attitude toward mathematics is not an artifact of aggregation and it cannot be explained psychologically. Evidently, there is a country factor, similar to the Jim Crow laws presented in the study by Subramanian et al. (2009), which controls both averages. Hypothetically, we may identify it with a primary school culture feature called pressure for knowledge acquisition.

In strong pressure systems, the curriculum determines the expected learning outcomes. Curriculum units (e.g., subjects of teaching) are carefully aligned so that each can contribute to the development of higher order competences set out as the aims of a teaching period (e.g., grade). Teaching is sequenced into meaningful units and its pace depends on pupils' progress in a way similar to Bloom's (1971) mastery learning. Virtually no pupil is allowed to move to the next unit unless there is evidence that he/she acquired the prescribed knowledge. Assessment is systematic, frequent, objective and nonnegotiable. There are final examinations at the end of school, often in the form of national or regional achievement tests.

In weak pressure systems, the curriculum outlines the teaching content and either links it loosely to autonomous subjects or abandons the subjects for "integrated" education. The aims of education are general and determined for the whole school period. The teaching pace is set out for the whole class. It is considered natural that some pupils make little or no progress at all. Assessment is subjective and open to dispute. It is interspersed by chance throughout the whole year. There is a tendency to avoid excessive probing into a pupil's mind, appreciate every scrap of knowledge and tolerate subtle evidence of ignorance. If a pupil fails, no specific steps are taken, except for blame. In short, strong pressure systems consider school to be a place of work, sometimes hard work, while in weak pressure systems, school is a place of gathering bits of knowledge with gay abandon. The latter was lucidly depicted in *Eugene Onegin* by Alexander Pushkin:

Мы все учились понемногу
Чему-нибудь и как-нибудь,
Так воспитаньем, слава богу,
У нас немудрено блеснуть.
Onegin was a prime example of such an education:
Он рыться не имел охоты
В хронологической пыли
Бытописания земли;
Но дней минувших анекдоты
От Ромула до наших дней
Хранил он в памяти своей[1].

[1] We've all acquired some education / A bit of this a bit of that,
God be thanked, some imitation, And we can all display éclat.
[…]
As for finding ancient treasure / He'd no desire to dig the dust
Of history all turned to rust, / But kept the juiciest stories ever
From Romulus to our own day, / In his memory tucked away.
(Translated by A.S. Kline, http://www.poetryintranslation.com/)

It is fairly obvious that strong country pressure for knowledge acquisition shifts the distribution of achievement to the right, while the distribution of attitude is shifted to the left, since high demands unavoidably put pupils under stress. The reverse holds for weak country pressure. At the international level, the pressure variable produces a negative correlation between achievement and attitude, as presented in Figure 1. Within countries, the variability of pressure for knowledge acquisition is restricted, hence pupils may maintain psychological coherence between achievement and attitude.

The concept of pressure for knowledge acquisition as a culturally specific feature of primary school education in a country helps to explain the pattern of findings reported above. The concept, however, needs independent substantiation through international research.

References

Alker, H.A. (1969). A typology of ecological fallacies. In: M. Dogan & S. Rokkan (eds.), *Quantitative ecological analysis* (pp. 69–86). Cambridge: MIT.

Bloom, B.S. (1971). Mastery learning. In: Block H.R. (ed.), *Mastery learning: Theory and practice*. New York: Holt, Rinehart and Winston.

Gove, W.R. & Hughes, M. (1980). Reexamining the ecological fallacy: A study in which aggregate data are critical in investigating the pathological effects of living alone. *Social Forces*, 58(4), 1157–1177.

Konarzewski, K. & Bulkowski, K. (red.) (2016). *TIMSS 2015. Wyniki międzynarodowego badania osiągnięć czwartoklasistów w matematyce i przyrodzie*. Warszawa: IBE.

Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Student Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351–357.

Selvin, H.C. (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, 63, 607–619.

Subramanian, S.V., Jones, K., Kaddour, A. & Krieger, N. (2009). Revisiting Robinson: The perils of individualistic and ecologic fallacy. *International Journal of Epidemiology*, *38*, 342–360.