

RESEARCH

Open Access

# What works where? The relationship between instructional variables and schools' mean scores in mathematics and science in low-, medium-, and high-achieving countries

Ruth Zuzovsky<sup>1,2</sup>

Correspondence: [ruthz@post.tau.ac.il](mailto:ruthz@post.tau.ac.il)

<sup>1</sup>Kibbutzim College of Technology,  
Education and the Arts, Ramat Aviv  
69978, Israel

<sup>2</sup>Science and Technology Center,  
School of Education, Tel Aviv  
University, Ramat Aviv 69978, Israel

## Abstract

**Background:** The association between frequent use of certain instructional practices in mathematics and science and learning outcomes in schools in low-, medium-, and high-achieving countries is the focus of this study. It not only looks at teaching practices "that work" (positively associated with achievement) but whether they "work" similarly in the three groups of countries.

**Method:** Hierarchical multilevel regressions analysis was employed to explore the relationships between frequent use of certain instructional practices in mathematics and science and schools' learning outcomes in these areas, in low-, medium-, and high-achieving countries.

**Results:** In both school subjects, traditional modes of instruction (teacher-centered) were found to be positively and significantly associated with achievement in all countries, while more constructive modes of instruction (student-centered) showed a differential effect. The frequent implementation of more student-centered modes was found to be positively associated with learning outcomes in high- and medium-achieving countries, but negatively associated in low-achieving countries.

**Conclusion:** The findings confirm conclusions in other studies that replacing teacher-centered traditional practices with more student-centered practices will not necessarily result in more learning for all students. Constructivist practices will be more beneficial for students only in high-achieving countries.

## Background

The decision to conduct cross-national comparative studies on the yield of educational systems by testing the achievements of comparable samples of students was reached in the late 1950s at the UNESCO Institute of Education in Hamburg, Germany. The International Association for the Evaluation of Educational Achievement (IEA) was established for this purpose. The founders of the IEA considered the idea of assessing the strengths and weaknesses of educational practices in a worldwide "educational laboratory" in which national educational policies and practices would be treated as inputs and student achievements and attitudes would be treated as outputs. It was expected that such a worldwide laboratory would make it possible to go beyond

descriptive identification of salient factors that account for cross-national differences toward explaining, predicting, and arriving at valid international generalizations regarding what works in education (Húsen, 1973).

Facing an educational reality in which variability exceeds similarity, the goal of constructing a comprehensive educational theory was relinquished, and instead of searching for similarities, researchers favored examining the differences that distinguished one country from another. Attempts to follow this line using data from all participating countries are rare. Researchers usually prefer to analyze data from only a few selected countries and delineate the differences or similarities among them (e.g., House, 2005; House & Telese, 2008; Le et al., 2006; Stevenson et al., 1987; Stigler et al. 1999). Among the studies that dealt with all participating countries are ones conducted by Schmidt et al. (1997a, b), Schmidt et al. (2001), and Houang et al. (2004), which aimed to identify differences and similarities that underlie the intended and implemented curricula of science and mathematics in all the countries that participated in the first cycle of Trends in International Mathematics and Science Study (TIMSS).

Other studies dealt with patterns of students' responses to test items in different countries (Dudits & Eljio, 2008; Grønmo et al. 2004; Rutkowski & Rutkowski, 2009; Zabulionis, 2001), which was defined as the "attained curricula." A different perspective on distinguishing among countries using the TIMSS 2003 database focused on clustering them according to teaching practices and attitudes toward mathematics as a school subject (Japeli-Pavesic & Korenjak-Cerne, 2004), or using TIMSS 1999 data to describe teaching practices in mathematics in 38 countries (Desimone et al. 2005).

All these studies focused mainly on differentiating among countries using descriptive measures of curriculum—implemented or attained, students' backgrounds and attitudes, classroom practices, school climate, students' responses to test items, and other contextual or outcome variables. The variability revealed in these studies led the researchers to conclude that teaching is not culturally independent (Fuller & Clarke, 1994). These conclusions were in line with findings drawn from other studies (Dale, 2000; Dale & Robertson, 2002), which provided evidence for regional similarities and argued for three regions of harmonized curricula and instruction: Europe, Asia, and America.

Adopting this view also directed me, in the early stages of the present study, to classify countries a priori according to cultural and geographical similarities, such as East Asia, Eastern Europe, or Arab countries, and so on, and to look for typical modes of instruction that characterize each of these groups of countries. The variability in the frequency of use of specific instructional practices within each group, together with the similarities in this regard in countries belonging to different groups, directed me, in the later stages of the study, to seek another classifying principle.

Instead of classifying the countries according to cultural or other independent-contextual variables (in this case, instructional practices), I chose to group the countries according to the dependent variables—the actual achievements of their students. The decision to turn to this type of classification was influenced by the methodology used in school effectiveness studies on "outlier schools"—those that achieve much more or much less than expected of them according to their student intake characteristics (Miller, 1985; Purkey & Smith, 1983). In those studies, characteristics of schools that "do well" are explored for their relationship with achievement. A similar approach was used by Postlethwaite and Ross (1992) in identifying variables that significantly discriminated between the 20% highest- and 20% lowest-

scoring schools that participated in TIMSS and PIRLS (Progress in International Reading Literacy Study).

Cutting the distribution of the countries' average achievement scores in mathematics and science on the TIMSS scale into three equally sized parts allowed me to define three groups of countries in terms of their performance: low-, medium-, and high-achieving. The three groups of countries created were found to be the same in both school subjects (see tables in Additional file 1: Appendix A). Each group comprised 15 to 17 countries and about 2,300 to 2,500 schools.

At this stage, the focus of the study also shifted from distinguishing among the groups of countries according to the frequency of use of instructional practices (the contextual variables) to distinguishing among them according to the size and type of relationship between these variables and the learning outcomes.

The research question phrased was, therefore, as follows: Are the relationships between frequent use of certain instructional practices and schools' learning outcomes in mathematics and science similar in the three groups of countries, or do they differ in the different groups? In other words, is there an interaction effect between the frequent use of certain instructional practices and the affiliation of schools with one of the three groups defined above that affects schools' outcomes? I was thus interested not only in delineating teaching practices "that work" (that are positively associated with achievement) but also in investigating whether they "work" similarly in all groups of countries. Hence, I am calling this paper, "What Works Where?"

Using data from international comparative studies provides, as Scheerens claims, an "interesting possibility to establish" whether what works in one country also works in the next. Stated in less popular terms, this question refers to the generalizability of "effectiveness enhancing conditions across countries" (Scheerens, 2004, p. 10).

### **Theoretical perspective**

The research question as phrased links this study to the research literature that deals with educational effectiveness or, more specifically, with instructional effectiveness. Most of the studies in this area fall under the category of process-product studies that deal with processes and conditions of teaching that enhance student outcomes (the product). In all studies that follow this line, the relationship between process variables and attainment is explored after controlling for student background variables.

The hypothesized relationship between processes of instruction and student attainment is rooted in different theoretical models on teaching and learning. Instructional modes that are backed by such theories are referred to as "instructional components" (Seidel & Shavelson, 2007, p. 456). Examples of such component are "Time on Task," a derivative of Carroll's (1963) model of teaching and learning; "Opportunity to Learn," a derivative of Bloom's (1976) model of mastery learning; or "Direct Teaching," borrowed from Doyle's (1985) model of teaching.

Review and meta/mega analyses carried out on the many studies of this type (Brophy & Good, 1986; Creemers, 1994; Fraser et al. 1987, Scheerens, 2000a, 2000b; Scheerens & Bosker, 1997; Scheerens, Vermeulen, & Pelgrun, 1989; Seidel & Shavelson, 2007; Stallings, 1985; Walberg, 1984; Wang et al. 1993) highlighted a number of such instructional components associated with achievement with highest effect sizes, such as time on task, structured

direct teaching, opportunity to learn, feedback and monitoring student progress procedures, and other variables that were later included in what Scheerens (1990) refers to as the "integrated model of school effectiveness."

The change that occurred in the last decade in learning and teaching theories, due to the new epistemological paradigm of constructivism, introduced new instructional components into the instructional effectiveness framework. These components focus more on students' active engagement in learning and construction of knowledge in real-world environments than on teachers' teaching behaviors. Among these new components, Seidel and Shavelson (2007) mention the following: "constructive learning," "domain specific," "social learning," "goal directed and self-regulated" and "evaluative" learning (for detailed description and references, see Seidel & Shavelson, 2007, p. 459–460).

The additional new component of instruction created a dichotomy between two types of instructional models. Scheerens (2004, p. 32) summarizes the differences between them. This appears in Table 1.

It should be emphasized that this bipolarity is not always accepted, and other scholars argue for more eclectic approaches and reconciliation between the two approaches (Brophy, 1996; Merrill, 1991).

Indeed, existing instructional practices in most school subjects represent a mixture of both traditional and constructivist instructional components. In mathematics there is a distinction between the "conceptual" model of instruction (in line with constructivist notions such as being engaged in problem-solving, working with real-world problems that have no obvious solutions, and discussing alternative solutions (Desimone et al., 2005; Hiebert et al., 1996) and the "computational" instructional approach that focuses on routine drill and practice and on traditional direct teaching (Li, 1999).

In science too, there is a distinction between traditional teacher-centered instructional practices (i.e., learning from textbooks, lectures, and memorizing scientific facts) and inquiry-oriented approaches (experimenting, problem solving using logic and evidence, elaborate explanations (Duschl, 1990; Shulman & Tamir, 1973; Von Secker, 2002; Von Secker & Lissitz, 1999).

In both subject areas, there is a debate as to whether the more constructivist approaches promote achievements of all students, or help only the brightest ones (Desimone et al., 2005; Le et al., 2006; Lee & Luyks, 2005; Tomlinson et al., 2003).

Data on instructional practices obtained from large-scale studies such as those carried out by IEA provided an opportunity to address such questions. Likert-type questionnaires developed in the context of IEA studies contained a list of statements describing typical modes of instruction in mathematics and science classes. These questionnaires were administered to teachers and their students. Teachers were asked how often students in their classes were engaged in different activities and students

**Table 1 Differences between traditional and constructivist models of instruction**

<b>Traditional instruction</b>	<b>Instruction inspired by constructivism</b>
Emphasis on basic skills	Bias toward higher order skills
Subject matter orientation	Emphasis on learning processes and cognition
Structured approach	Discovery learning
Abstract-generalized knowledge	Situation specific knowledge
Standardized achievement tests	Performance assessment

responded on the same scale on the frequency of being exposed to the various modes of instruction.

The statements in the questionnaires reflected the reality that exists in classrooms and the state of art concerning instruction in the two subject areas. They were carefully phrased by content and psychometric experts in a way that will allow them to be used in different countries. The questionnaires were then field tested, revised, and modified over the years in order to be tuned in to changes in the ways mathematics and science are taught (a full description of the process of developing the questionnaires is in Erberber et al. 2008; Arora et al. 2004).

The teaching and learning activities that were addressed in the questionnaires represented a mix of traditional and constructivist learning and teaching activities. Examples from mathematics are traditional activities such as “listen to teachers giving a lecture-style presentation,” and “memorize facts and procedures,” or statements that represent constructivist modes of instruction, that is, “work out problems on our own,” “relate what is learned to daily life,” “decide on our own on procedures for solving complex problems,” and so on.

The opportunity to assess the effectiveness of these instructional variables in a multitude of countries using these questionnaires brings us back to the present study.

## **Method**

The data that served this study were obtained from the TIMSS 2007 database. For each of the 49 countries that participated, it provided estimated proficiency (achievement) scores in mathematics and science on the TIMSS scores scale, with an average score set to 500 and a standard deviation to 100, and extensive data on contextual variables—social as well as educational. TIMSS scaling approach uses multiple imputation or “plausible values” methodology to obtain proficiency scores in each subject area for the entire population.

Hierarchical multilevel regression analysis using Hierarchical Linear and Nonlinear Modeling HLM.6 software (Raudenbush et al. 2004), was employed to explore the relationship between the frequency of using a set of instructional variables and the average score of schools in mathematics and science.

The models specified for this analysis were two-level models of schools (7,347) nested in countries (49) that participated in TIMSS 2007. Because of missing data, the data that served the HLM analyses in this study represented only 7,201 schools from 48 countries. As the TIMSS sample design allowed sampling one class in each sampled school, data that were obtained on the class level also represented the school level.

The school (class) level was decided upon as the appropriate lower level of the analysis as this is the level where our target variables—the instructional practices—operate and the aim of the study as defined was to explore the association of their frequent use with the average score of the school. The purpose was to look for this association only on the class/school level.

The analyses reported here are based on class/school averages of five imputed plausible values for each subject area. The plausible values reflect the need to impute student performance on the entire item pool from their performance on only a subset of items they took as occurs in TIMSS studies. Differences across plausible values thus reflect

the uncertainty associated with the measurement of the proficiency variable. The choice made to use the average of all five plausible values was meant to ease the computational burden of the analyses. Consequently the standard errors do not reflect the imputation uncertainty and so underestimate the full level of uncertainty.

Due to the effect that student body composition might have on such an association, important student background variables aggregated at the class/school level were used to control for their effect at this level of analysis. Thus, on the school level, the specified models included two aggregated student level variables that described students' background: Aspiration to complete higher levels of education (HFG) on a scale of 1 (finish secondary school) to 5 (beyond first university degree); and the number of books at home (book)—a proxy for students' sociocultural background—on a scale of 1 (few) to 5 (many). Another variable specified on this level described our target variable—the school (class) mean of students' perceptions of the frequency of being exposed to one of several modes of instruction on a scale of 1 (in every, or almost every, lesson) to 4 (never). As all instructional variables are on a scale from 1 to 4, there was no need to center them in the multilevel analysis.

On the country level, dummy variables were used to indicate the schools' affiliation to one of the three equally sized groups of countries established: (1) low-, (2) medium-, and (3) high-achieving groups of countries. The medium-achieving group was chosen to serve as the comparison group to which estimated regression coefficients for high- and low-achieving countries were compared.

In addition to the null model that was used to partition the total variance of schools' average scores in science or mathematics to "between schools" and "between countries" components, and a model that included only school-aggregated student background variables, three alternative explanatory models for each of the instructional variables in the two school subjects were specified. The first model included, in addition to the school-aggregated student background variables, the school-aggregated students' perceptions of the frequency of implementing one specific instructional mode. (There were 17 instructional modes in mathematics and 16 in science.) In the second model, dummy variables indicating the schools' group affiliation were specified for Level 2. The two dummy variables compared the low- and the high-achieving countries with the middle achievement group. The third model also included the interaction terms between the relevant instructional variables and the two dummy variables on Level 2.

This last model was meant to provide an answer to the research question: Is the association between frequent use of certain instructional practices and the mean achievement score of schools in mathematics and science similar in all three groups of countries or does it differ from high- to mid- to low-achieving countries? The regression equation of the model with the interaction terms is the following:

#### Level 1 Model

$$Y = B_0 + B_1*(HFG) + B_2*(BOOK) + B_3*(Relevant Instructional Variable) + R$$

#### Level 2 Model

$$B_0 = G_0 + G_{01}*(Group 1) + G_{02}*(Group 3) + U_0$$

$$B_1 = G_1 + U_1$$

$$B_2 = G_2 + U_2$$

$$B_3 = G_3 + G_{31}*(Group 1) + G_{32}*(Group 3) + U_3$$

The regression coefficient (B) of the relevant instructional variable obtained from the third model indicated the size and direction of change in the mean achievement score of schools as a result of a one-unit change on the frequency scale of implementing that instructional variable in the medium-achieving group of countries.

The regression coefficient of the interaction term between the frequent use of instructional variable and the school affiliation to either the low- or high-achieving group of countries indicates the change in the mean achievement score of schools as a result of a one-unit change on the frequency scale of using the relevant instructional mode in these two groups of countries, as compared to such a change in the medium-achieving group. These interaction term coefficients, when added to the regression coefficient of the instructional variables in the medium-achievement group, provide us with the regression coefficient of the instructional variable in the low- and high-achievement groups.

The values of the regression coefficient of the relevant instructional variable in the medium-achieving group of countries as well as the values of the regression coefficient of the interaction term between the instructional variable and the variables that indicate school affiliation to the low- and high-achieving countries represent a cardinal outcome of the analyses and are reported later in the results section.

Specifying the comparison group several times, each time using another group of countries, allowed me to obtain not only the size and direction of the regression coefficient of the relevant instructional variable in each group, but also its statistical significance.

Given the large number of interaction coefficients that were derived from running three models for each of the instructional variables (16 and 17 variables in the two subject areas) with two interaction terms for each of them, caution is needed in interpreting the results due to the increased probability of Type I error. As the study is exploratory in nature, the regression coefficients appearing in Tables 2 and 3 are not corrected for multiple comparisons. Many of the coefficients appearing in the tables still do not reach statistical significance. Employing the Bonferroni procedure for "multiple comparisons" further reduces the number of statistically significant regression coefficients, and the reader should consider only those coefficients appearing in the tables with a significance level of  $p \leq .000$  as statistically significant at the 0.05 level.

In this regard, I prefer to assess the meaning of the regression coefficient by comparing its size to the standard deviation of the distribution of the schools' mathematics and science mean scores in each group of countries. In mathematics, the standard deviation of this distribution was 59 points in low-achieving countries, 60 in medium-achieving countries, and 63 in high-achieving countries. In science, the parallel standard deviations were 61 in low-achieving countries, 54 in medium-achieving countries, and 51 in high-achieving countries.

Regarding students' academic aspirations, and the number of books in students' homes, a positive regression coefficient indicates a positive relationship with the average achievement score of the school. In the case of the frequency of the instructional mode, on a scale from 1 (very frequent) to 4 (never), a negative regression coefficient indicates a positive association with the average achievement score of the school.

The association between the frequency of implementing an instructional mode and the average achievement score of schools in each group of countries can be visualized

**Table 2 Regression coefficients of instructional variables on schools' mean scores in mathematics**

Instructional practice	Regression coefficient		
	Low-achieving countries	Medium-achieving countries	High-achieving countries
<i>(a) Instructional practices that focus on developing computational skills</i>			
Practice four operations without using a calculator (ASAM)	-11.4* (13.4)*	-24.8**	-1.0 (23.8**)
Write equations and fractions (EFR)	-7.8* (3.6)	-11.4**	-14.3* (-2.9)
Memorize formulas and procedures (FRR)	-19.5*** (-1.9)	-17.6*	-8.9 (8.7)
Solve problems in geometry (GSA)	2.3 (7.1)	-4.8	-3.8 (1.0)
Interpret data in tables, charts and graphs (GCT)	14.6* (7.5)	7.1	11.4* (4.3)
Work on fractions and decimals (WFD)	22.2*** (30.0***)	-7.8	-3.0 (4.8)
<i>(b) Traditional practices</i>			
Review homework (ROH)	-15.4* (-4.3)	-11.1*	-5.7* (5.3)
Listen to teacher lecture (LSP)	-10.8 (7.4)	-18.2*	-12.7* (5.5)
Begin homework in class (BHC)	27.4*** (-22.9***)	4.4	5.3 (0.9)
Have quizzes or tests (HQT)	18.7** (7.8)	10.9	4.8 (-6.1)
Use computers (COM)	41.5*** (18.3**)	23.1***	16.9*** (-6.3)
Use calculators (CAL)	4.6 (-5.9)	10.5*	-6.1 (-16.6*)
<i>(c) Constructivist modes of instruction</i>			
Students work out problems on their own (WPO)	12.7** (44.5***)	-31.8***	-28.5*** (3.3)
Students explain their answers (EXP)	-26.1** (6.0)	-32.0***	-14.3** (17.8)
Students decide on their own about procedures for solving problems (SCP)	11.0* (14.7)	-3.6	-12.7* (-9.0)
Work together in small groups (WSG)	15.1** (-0.3)	15.4**	11.1*** (-4.3)
Relate what is learned to daily life (MDL)	9.2* (5.5)	3.7	1.9 (-1.8)

Note: In parentheses – regression coefficients of the instructional variable with school's affiliation with either low- or high-achieving groups of countries, compared to that in the medium-achieving group.  
 \*\*\* =  $p \leq .000$ ; \*\*  $p \leq .01$ ; \*  $p \leq .05$ .

by plotting a line on a graph between the predicted schools' mean score in mathematics or science at two distal frequencies of implementing the instructional mode. (See an example of such a prediction in Additional file 2: Appendix B.)

As the frequency scale runs from 1 (very frequent) to 4 (never), an upward line indicates a negative association between implementing the instructional mode and the schools' mean score, and a downward line indicates a positive association.

An additional outcome of the multilevel regression analysis was information on the variance components of the between-schools and between-countries average school score in mathematics and in science (details presented in the Results section).

## Results

### Instructional modes in mathematics

Student perception of the frequency of using 17 modes of instruction common in mathematics classrooms were aggregated on the school level. These practices

**Table 3 Regression coefficients of instructional variables on schools' mean scores in science**

Instructional practice	Regression coefficient		
	Low-achieving countries	Medium-achieving countries	High-achieving countries
<i>(a) Traditional teacher-led instructional practices</i>			
Listen to teacher lecture (LSP)	<b>-17.9*</b> (4.2)*	<b>-22.11**</b>	<b>-25.6***</b> (-3.5)
Watch teacher demonstrate an experiment (DEI)	<b>-12.5*</b> (-5.3)	<b>-7.1*</b>	<b>-0.1*</b> (7.0)
Memorize facts and principles (FAP)	<b>-26.3***</b> (-6.4)	<b>-19.8**</b>	<b>-15.6**</b> (4.2)
Use scientific formulas and laws to solve problems (LAW)	<b>-8.2</b> (6.8)*	<b>-15.0*</b>	<b>-13.4*</b> (1.6)
Review homework (ROH)	<b>-12.4</b> (-4.1)*	<b>-8.4*</b>	<b>-4.9</b> (3.5)
Begin homework in class (BHC)	<b>36.0***</b> (15.4)	<b>20.6**</b>	<b>11.1**</b> (-9.6)
Read science textbooks and other resource materials (TEX)	<b>17.7*</b> (33.0**)	<b>-15.3*</b>	<b>-15.9***</b> (-0.6)
Have quizzes or tests (HQT)	<b>16.0**</b> (10.4)	<b>5.6*</b>	<b>11.3*</b> (5.7)
Use computers (COM)	<b>48.3***</b> (32.9***)	<b>15.4**</b>	<b>7.4**</b> (-8.0)
<i>(b) Inquiry-oriented, constructivist student-centered practices</i>			
Make observations and describe what is seen (OBS)	<b>6.2</b> (8.8)	<b>-2.6</b>	<b>-6.4*</b> (-3.8)
Conduct experiments (CEI)	<b>3.8</b> (7.9)	<b>-4.2</b>	<b>-6.9*</b> (-2.7)
Design or plan experiments (PEI)	<b>10.4**</b> (4.2)	<b>6.2</b>	<b>3.4</b> (-2.8)
Work in small groups on an experiment or investigation (WGO)	<b>3.6</b> (6.3))	<b>-2.7</b>	<b>-6.3</b> (-3.6)
Give explanation about what is studied (EOS)	<b>-16.5</b> (8.0))	<b>-24.5***</b>	<b>-16.8***</b> (7.7)
Work on problems on their own (WPO)	<b>22.6**</b> (22.3*)	<b>0.4</b>	<b>-9.8**</b> (-10.2)
Relate learning to daily life (MDL)	<b>-7.6</b> (5.5)	<b>-13.1*</b>	<b>-8.6*</b> (4.5*)

Note: In parentheses – regression coefficients of the instructional variable with school's affiliation with either low- or high-achieving groups of countries, compared to that in the medium-achieving group.  
 \*\*\* =  $p \leq .00$ ; \*\* =  $p \leq .01$ ; \* =  $p \leq .05$ .

were classified into three groups: those that focus on developing computational skills; those that represent traditional, mostly teacher-led, instructional practices; and those that represent conceptual, more constructivist practices. This classification echoes the distinction between traditional and constructivist modes of instruction discussed earlier in this paper. The descriptive statistics of the frequency of using these variables as well as their regression coefficients follow this classification.

The descriptive statistics of the frequency of using mathematics instructional modes are in Additional file 3: Appendix C. In the following section, data on the regression coefficients are presented.

***The relationship between frequent use of mathematics instructional modes and mathematics achievement***

Table 2 shows the regression coefficient of the different instructional modes used in mathematics classrooms on schools' mathematics mean scores in low-, medium-, and high-achieving countries (in bold in the table). The values appearing in the table were obtained from separate analyses carried out for each instructional variable. The size and sign of the coefficients indicate the strength and direction of the association between frequency of using a specific instructional mode and the average mathematics score in schools. The table also presents the regression coefficient of the interaction

terms (in brackets) between the instructional variables and the affiliation of schools to either the low- or high-achieving group of countries.

### ***Interpretation***

Frequent use of three of the six instructional modes that aim at developing computational skills were found to be positively associated with the mean mathematics score of schools in all three groups of countries, although in varying strength. For two modes, "memorizing formulas and procedures" and "practicing the four arithmetical operations without using a calculator," this association is more profound in low- and medium-achieving countries, while in the case of the third variable, "writing equations and functions to represent relationships," the association is more profound in high- and medium-achieving countries. A one-unit increase on the frequency scale of these variables results in an increase of about 0.15 to 0.4 standard deviations of the distribution of schools' mean scores in the different groups of countries. Frequent use of other instructional modes of this type ("working on fractions and decimals," "interpreting data in tables, charts and graphs") were found to be negatively associated with achievement in low-performing countries.

In the case of traditional teacher-led modes of instruction, such as "reviewing homework" and "listening to teacher lecturing," a one-unit increase on their frequent use scale increases the mathematics mean scores by about 0.1 to 0.3 of the standard deviation of the distribution of schools' mean scores in the relevant group of countries. The positive association of "reviewing homework" with achievement is more profound in low- and medium-achieving countries, while that of "listening to teacher lecturing" is more profound in medium- and high-achieving countries. Some traditional activities, such as "having quizzes or tests" or "beginning to do homework in class," when occurring frequently, were found to be negatively associated with the average achievement of schools in all groups of countries. Frequent use of computers was also found to have a negative association with average mathematics achievement of schools in all groups of countries.

Among the more constructivist modes of instruction, only the practice that requires students to "explain their answers" was found to be highly associated with the average achievement of schools in all three groups of countries; more so in medium- and low-achieving countries. However, requiring students to "work out problems on their own" or to "decide on their own on procedures for solving complex problems" was found to be positively associated with the average achievement of schools in high- and medium-achieving countries, but negatively associated with achievement in low-achieving countries.

Generally speaking, it seems that practices that focus on computational skills and traditional teacher-led, more direct instruction that are positively associated with achievement in all groups of countries are more effective in low- or medium-achieving countries, while more challenging constructivist modes of instruction are more effective in medium- or high-achieving countries.

Plotting graphs of predicted schools' mean mathematics scores at two distal categories on the frequency scale of using these practices (1 and 4) makes it possible to visualize the association between frequent use of instructional practices and the schools' mean mathematics scores. Some practices, such as "listen to teacher lecture" or

"students explain their answers," are positively associated with mathematics achievement in all groups of countries. Others, for example "work together in small groups," exhibit a negative association in all groups of countries. Differential association occurs, for example, when students are asked to "work out problems on their own" or "decide on their own on procedures for solving problems."

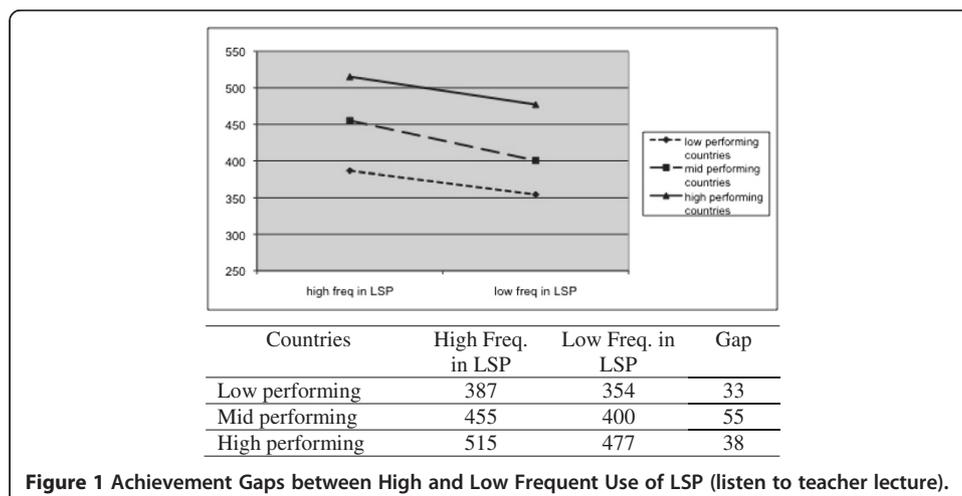
To give a feeling of these patterns of association, three plots are presented. Figure 1 illustrates the positive relationship between a traditional teacher-centered mode of instruction – LSP ("listen to teacher lecture") and schools' mean score in mathematics, while Figure 2 illustrates a negative association, this time between another common practice, "working in small groups" (WSG) and schools' mean mathematics scores.

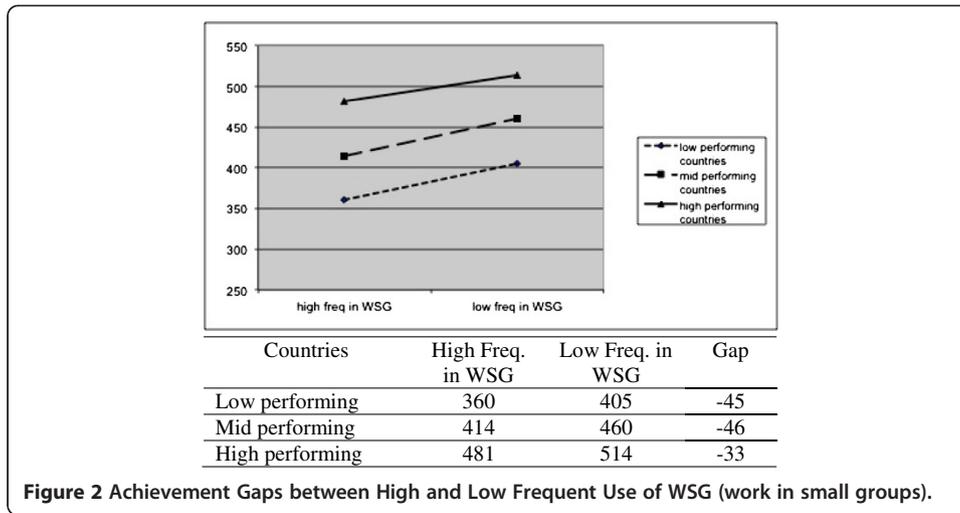
Figure 3 demonstrates the differential relationship between "students working out problems on their own," a practice that reflects more student-centered constructivist modes of instruction, and schools' mathematics mean scores. Here the relationship is positive in high- and medium-achieving countries and negative in low-achieving countries.

Predicted scores for each group of countries at the two distal categories on the frequency of use scale of the instructional practice make it possible to calculate the achievement gap between students exposed frequently to the practice and those who never engage in such practices. These gaps, when compared with the standard deviation of schools' mathematics mean scores, are considered large.

### Instructional modes in science

Since the mid-20th century, traditional expository teacher-led instruction in science has given way to inquiry or discovery modes of learning. Advocated by Bruner (1961) and Schwab (1962), the concepts of the "structure of the disciplines" and the slogan "learning science as inquiry" shaped science curricula and learning practices in many classrooms around the world. In more updated versions of inquiry learning, constructivist notions of learning, performed by the students themselves or with the scaffolding of others, appeared. Following such a distinction, instructional practices in science classrooms were classified into two groups of practices: traditional teacher-led



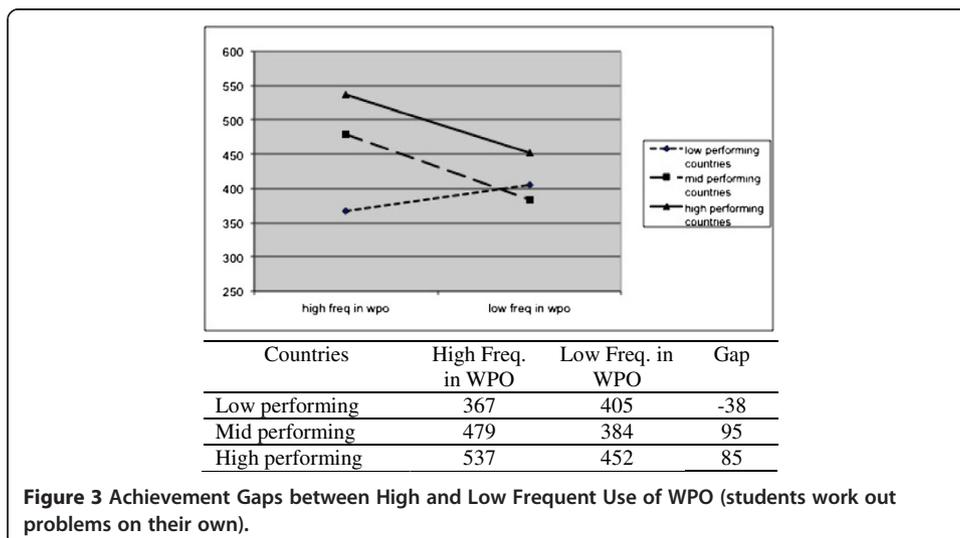


**Figure 2** Achievement Gaps between High and Low Frequent Use of WSG (work in small groups).

expository practices and the more inquiry-oriented, student-led, constructivist modes of instruction. Here, too, this classification echoes the "instructional components" that appear in school effectiveness literature discussed earlier. Descriptive statistics on the frequency of using these practices as well as their regression coefficients on the average science score of schools in the three groups of countries follow this classification. The descriptive statistics of the frequency of using the instructional practices appear in Additional file 4: Appendix D. In the following section, the regression coefficients of these variables are presented.

***The relationship between frequent Use of science instructional modes and achievement in science***

Table 3 shows the regression coefficients of different practices in the low-, medium- and high-achieving groups of countries (in bold in the table). As in Table 2, this table presents the regression coefficient of the interaction terms (appearing in brackets) between the instructional variables and the variable that indicates the affiliation of the schools to either the low- or the high-achieving group of countries.



**Figure 3** Achievement Gaps between High and Low Frequent Use of WPO (students work out problems on their own).

### ***Interpretation***

Despite the advocacy for inquiry-oriented student-centered modes of instruction, such as "make observations," "plan and conduct experiments," "work on experiments in small groups," and so on, the regression coefficients of such practices, even if showing a positive association with the average science score of the school, are small and in most cases statistically insignificant. The only student-led activity found to be positively and significantly associated with science achievement in schools in all groups of countries was that of students "providing their own explanations about what they study." A one-unit increase on the frequency scale of this activity increases the average school score in all groups of countries from about 0.27 to 0.45 of a standard deviation of the distribution of schools' mean scores in the relevant groups of countries. Another practice associated with constructivist notions of learning and recently used in science classrooms, "relating what is learned to daily life," also seems to be positively associated with achievement in all groups of countries. However, this association is weak. A one-unit increase on the frequency scale of this activity increases average school scores only from 0.1 to 0.24 of a standard deviation of the mean distribution of school scores in the relevant group of countries.

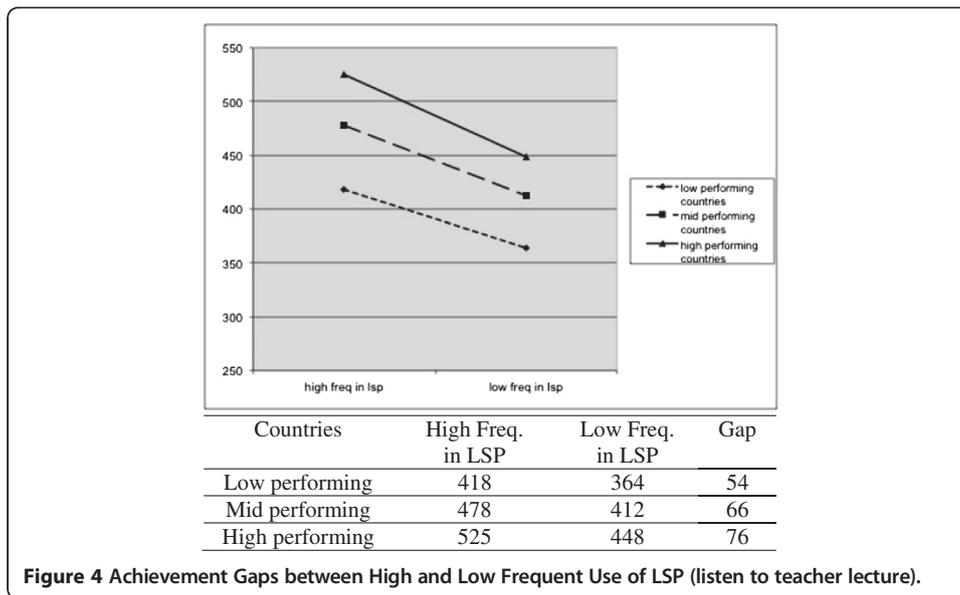
In contrast to the unfulfilled expectations of inquiry-oriented and constructivist modes of instruction, many traditional teacher-led practices in science classrooms, such as "listening to teacher lecturing," "memorizing facts and principles," "using formulas and laws to solve problems," "reading textbooks," and so on, were found to be positively associated with the mean science scores of schools in all groups of countries.

Some traditional practices, when frequently implemented, are more positively associated with the mean score of schools in low-achieving countries. Such is the case when students are often asked to "memorize science facts and principles" or to observe their "teacher demonstrating an experiment," but in other cases, such as "listening to teacher lecturing," "using scientific formulas and laws to solve problems," or "reading textbooks and other resource material," this positive association is more profound in high- and medium-achieving countries.

It is interesting to note that while frequent "reading of textbooks" is positively associated with the average science scores in schools in high- and medium-achieving countries, it is negatively associated with the average science scores in schools in low-achieving countries.

Some traditional practices, such as "beginning homework in class," "having quizzes or tests," and "using computers," when frequent, were found to be negatively and significantly associated with achievement in schools in all groups of countries.

Plotting a graph of predicted school mean science scores for the distal categories on the frequency scale of using some instructional variables shows that there are some that "work" similarly in all groups of countries, although with varying strengths, while others exhibit a differential effect in the different groups of countries. "Listening to teacher lecturing," "memorizing science facts and principles," and "giving explanations" are practices which have a positive association with school mean science scores in all groups of countries, while "working independently on solving problems" and "reading textbooks and other resource material" are practices that have a differential effect. Three examples are shown below to demonstrate the patterns of association.

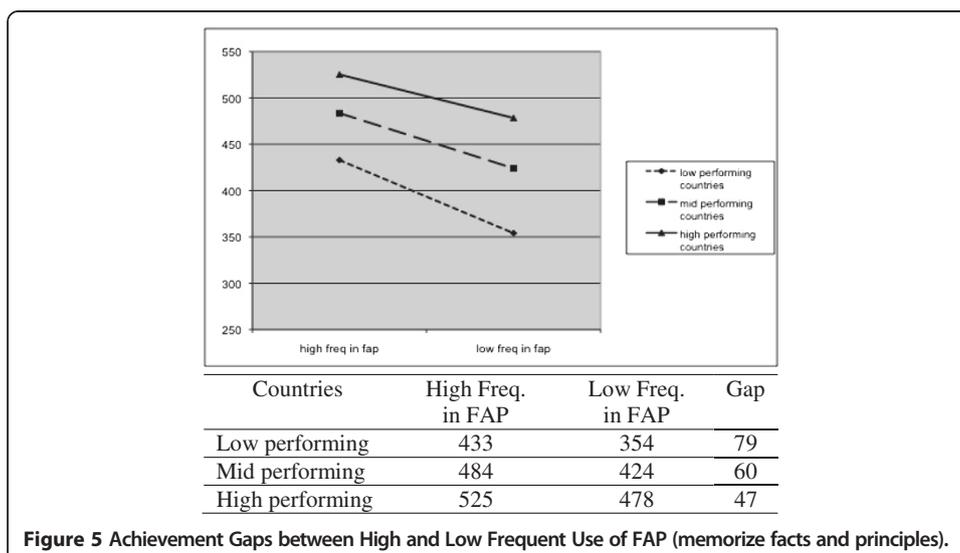


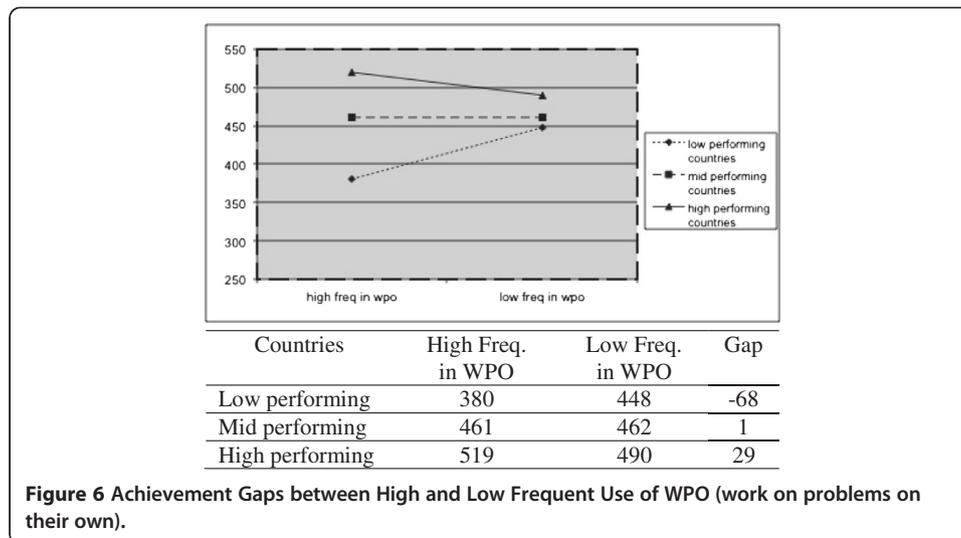
Figures 4 and 5 show a positive association between "listening to teacher lecturing" and "memorizing facts and principles" and the mean science score of schools. In contrast, Figure 6 displays the differential association between "students working out problems on their own" and the mean science score of schools.

Predicted scores in the two distal categories on the frequency scales of using the instructional practice make it possible to compute the achievement gap between students exposed during every lesson to the instructional practice and those who never engage in such a practice.

***The between-schools and between-countries variance components***

The hierarchical regression analysis employed in this study yields information on the variance components of the between-schools and between-countries average school scores in mathematics and in science. The variance components of the average school





mathematics scores are 2,683 between-schools and 4,952 between-countries – 35% vs. 65% of the total variance in average school scores.

In science, the variance components of average school science scores are 2,257 between schools vs. 3,326 between countries—43% vs. 57% of the total variance in average school scores. While most of the variances in schools' average mathematics scores lies between countries, in science the difference between the variance components is less pronounced.

### Conclusion

The regression coefficient of a set of instructional variables on the mean score of schools in mathematics and science in the three groups of countries provides us with an answer to the research question of whether or not frequent use of these modes of instruction is similarly and significantly associated with learning outcomes in all three groups of countries. A similar association might support the idea of an existing comprehensive instructional theory about "what works" and "what does not work."

Indeed, in both school subjects, some modes of instruction were found to be similarly associated with achievements in all three groups of countries (either positively or negatively). In mathematics, instruction targeted at developing computational skills (practicing four operations without calculators, memorizing formulas and procedures, and writing equations and functions) and traditional modes of instruction (listening to teacher lecturing, and requiring students to explain their answers), were found to be positively and significantly associated with mathematics achievements in all groups of countries, though with varying strength. Usually, this association is much stronger in low-achieving countries.

Some instructional activities in mathematics classes were found to be negatively and significantly associated with achievement in all groups of countries and often more so in low-achieving countries (frequent interpretation of data in tables or graphs, begin homework in class, frequent use of computers, frequent group work, and having tests or quizzes frequently); while interpreting graphs and charts, and the use of computers are regarded as more demanding modes of instruction. Frequent testing or frequent group work and frequently starting to do homework in class may be a symptom of low attainment and not its cause.

In science, too, certain instructional variables were found to be similarly associated with achievement in all groups of countries. As in mathematics, variables that represent traditional expository modes of teaching (listening to teacher lecturing, and memorizing facts and principles) were found to be positively and significantly associated with science achievements in all three groups of countries. Here too, some types of instruction were found to be negatively associated with achievements in all groups of countries (frequent use of computers, frequent testing, frequently beginning to do homework in class). As in the case of mathematics instruction, the negative association of frequent use of computers with achievement can signal teachers' lack of digital pedagogies, which might explain the ineffectiveness of this mode of instruction. The habit of starting to do homework in class may indicate reduced instructional time, which may have a negative effect on achievement, or may signify weakness of the students that is, on its own, the reason for their low achievements. Similarly, the negative association of frequent testing with achievement may be the result of low performance of schools and not its cause.

On the other hand, in both school subjects, there are variables that do not exhibit similar association with achievement in all groups of countries. Most of these variables are those oriented toward more constructivist modes of instruction. Variables describing students working on problems on their own both in science and mathematics—designing or planning experiments in science, and deciding on ways to solve problems in mathematics, which are highly demanding modes of instruction—show a differential effect. Frequent implementation of these practices was found to be positively associated with learning outcomes in high- and medium-achieving countries but negatively associated with learning outcomes in low-achieving countries.

These findings confirm conclusions reached in other studies of science and mathematics instruction (Le et al., 2006; von Secker, 2002; von Secker & Lissitz, 1999), which hold that replacing teacher-led (traditional) practices with more student-led (constructivist) practices will not necessarily result in more learning for all, unless students have the basic vocabulary and conceptual understanding essential for engaging in meaningful self-regulated learning. Such student-centered, more demanding practices will be more beneficial for high-achieving students (more of whom can be found in high-achieving countries) and might be a waste of time for low-achieving students in low-achieving countries.

The association between modes of instruction and students' academic achievements observed in this study on the school level provide clues regarding the different approaches to teaching mathematics and science that are sensitive to the students' academic level. The instructional choices that teachers make do not affect all students in all countries equally. Evaluating the differential effect of teacher practices in different countries grouped according to achievement level can help to shape effective pedagogical practices and also have implications for teacher training in different countries.

#### **Limits of the study**

As this study follows the tradition of instructional effectiveness studies and uses advanced statistical procedures employed in the area, it cannot escape the significant criticism raised against this line of research (Caro & Sandoval-Hernandez, 2012a; Sandoval-Hernandez, 2008; Wrigley, 2004). The most common critique points to the lack of theory behind school effectiveness studies. Some claim that the operationalization of the constructs included in these studies represents "no more than common sense and statistical criteria

without considering the theories available in education and other disciplines." If such constructs turn to show significant coefficients in statistical models, they are regarded as important factors for improving education systems (Caro & Sandoval-Hernandez, 2012b, p. 2). This kind of wrong causal inference, normally coupled with additive causal interpretation, ignores the complex nature of educational systems in which educational outcomes result from interactions (Aitkin & Zuzovsky, 1994; Murnane & Willett, 2011).

A related criticism concerns the "fishing for correlations" practice between particular constructs and student outcomes without fully understanding why and how it is expected that the two would be related (Coe & Fitz-Gibbon, 1998). On this aspect, Caro and Sandoval-Hernandez (2012b) note that it can be said that we know something about what works in education, but we know little about why it works or about the mechanisms at work.

This study, thus, is a challenge for further studies to provide answers to the "why" questions.

A further limitation relates to the use of classroom averages, thus ignoring the imputation variance of the plausible values. Readers should take into consideration that this analysis choice will deflate standard errors.

### Additional files

**Additional file 1: Appendix A.** Countries grouped by achievement in science and mathematics.

**Additional file 2: Appendix B.** Predicted score for high and low frequencies of WPO.

**Additional file 3: Appendix C.** Descriptive statistics of the frequency of using instructional variables in mathematics in the three groups of countries.

**Additional file 4: Appendix D.** Descriptive statistics of the frequency of using instructional variables in science in the three groups of countries.

### Competing interests

The author declares she has no competing interests.

Received: 9 May 2013 Accepted: 20 May 2013

Published: 9 July 2013

### References

- Arora, A, & Ramirez, MJ. (2004). Developing indicators of education contexts. In Constantinos Papanastasiou (Ed.), *TIMSS Proceedings of the IRC-2004 Conference* (Vol 1, pp. 1–18). Cyprus University Press.
- Aitkin, M, & Zuzovsky, R. (1994). Multilevel interaction models and their use in the analysis of large-scale school effectiveness studies. *School Effectiveness and School Improvement*, 5(1), 45–73.
- Bloom, B. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Brophy, J. (1996). *Classroom management as socializing students into clearly articulated roles*. Paper presented at the AERA Annual Meeting. New York: Meeting.
- Brophy, J, & Good, TL. (1986). Teacher behavior and student achievement. In M Wittrock (Ed.), *Third Handbook of Research on Teaching* (pp. 328–375). New York: Macmillan.
- Bruner, JS. (1961). The act of discovery. *Harvard Educational Review*, 31, 21–32.
- Caro, D, & Sandoval-Hernandez, A. (2012a). *An exploratory structural equation modeling approach to evaluate sociological theories in international large-scale assessment studies*. Paper presented at the AERA Annual Meeting. Canada: Vancouver.
- Caro, D, & Sandoval-Hernandez, A. (2012b). *An exploratory structural equation modeling approach to evaluate sociological theories in international large-scale assessment studies*. Paper presented at the AERA Conference. Canada: Vancouver.
- Carroll, JB. (1963). A model of school learning. *Teachers College Record*, 64, 722–733.
- Coe, R, & Fitz-Gibbon, CT. (1998). School effectiveness research: Criticisms and recommendations. *Oxford Review of Education*, 24(4), 421–438.
- Creemers, BPM. (1994). *The effective classroom*. London: Cassell.
- Dale, R. (2000). Globalization and education: Demonstrating a "common world educational culture" or locating "structural education agenda"? *Educational Theory*, 50(4), 427–448.
- Dale, R, & Robertson, SL. (2002). The varying effects of regional organizations as subjects of globalization of education. *Comparative Education Review*, 46(4), 10–36.
- Desimone, LM, Smith, T, Baker, D, & Ueno, K. (2005). Assessing barriers to reform of US mathematics instruction from an international perspective. *American Educational Research Journal*, 42, 501–536.
- Doyle, W. (1985). Effective secondary classroom practices. In MJ Kyle (Ed.), *Reaching for excellence. An effective schools' sourcebook*. Washington, DC: US Government Printing Office.

- Dudits, J, & Eljijo, A. (2008). *Trends in similarities and differences of students' mathematics profiles in various countries. Paper presented at the Third IEA Research Conference*. Taipei: The National Taiwan Normal University.
- Duschl, RA. (1990). *Restructuring science education: The importance of theories and their development*. New York: Teachers College Press.
- Erberber, E, Arora, A, & Preuscheft, C. (2008). Developing the TIMSS-2007 background questionnaires, *TIMSS-2007 Technical Report*. In *TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College* (pp. 45–62).
- Fraser, BJ, Walberg, HJ, Welch, WW, & Hattie, JA. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 145–252.
- Fuller, B, & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules and pedagogy. *Review of Educational Research*, 64, 119–157.
- Gronmo, LS, Kjaernsli, M, & Lie, S. (2004). Looking for cultural and geographical factors in patterns of responses to TIMSS items. In EC Papanastasiou (Ed.), *Proceedings of the IRC-2004 TIMSS Conference* (pp. 99–112). Cyprus: Cyprus University Press.
- Hiebert, J, Carpenter, TP, Fennema, E, Fuson, K, Human, P, Murray, H, et al. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, 25(4), 12–21.
- Houang, RT, Schmidt, WH, & Cogan, L. (2004). Curriculum and learning gains in mathematics: Across country analysis using TIMSS. In C Papanastasiou (Ed.), *Proceedings of the IRC-2004 TIMSS Conference* (Vol. 1, pp. 224–254). Cyprus: Cyprus University Press.
- House, JD. (2005). Motivational qualities of instructional strategies and computer use for mathematics teaching in the United States and Japan: Results from TIMSS-1999 – assessment. *International Journal of Instructional Media*, 32, 89–104.
- House, JD, & Telese, J. (2008). Relationships between student and instructional factors and algebra achievement of students in the United States and Japan: An analysis of TIMSS-2003 data. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14(1), 101–112.
- Hüsen, T. (1973). Foreword. In LC Comber & JP Keeves (Eds.), *Science education in nineteen countries. International Studies in Evaluation* (Vol. 1). Stockholm: Almquist & Wiksell.
- Japeli-Pavesic, B, & Korenjak-Cerne, S. (2004). Differences in teaching and learning mathematics in classes over the world: The application of adapted leader clustering method. In EC Papanastasiou (Ed.), *Proceedings of the IRC-2004 TIMSS Conference* (Vol. 2, pp. 85–107). Cyprus: Cyprus University Press.
- Le, V, Stecher, BM, Lockwood, JR, Hamilton, LS, Robyn, A, Williams, VL, Ryan, GW, Kerr, KA, Martinez, JF, & Klein, SP. (2006). *Improving mathematics and science education: A longitudinal investigation of the relationship between reform-oriented instruction and student achievement*. Santa Monica, CA: Rand.
- Lee, O, & Luyks, A. (2005). Dilemmas in scaling up innovations in elementary science instruction with non-mainstream students. *American Educational Research Journal*, 42, 411–438.
- Li, S. (1999). Does practice make perfect? *For the Learning of Mathematics*, 19(3), 33–35.
- Merrill, MD. (1991). Constructivism and instruction design. *Educational Technology*, 31, 45–53.
- Miller, SK. (1985). Research on exemplary schools: An historical perspective. In G Austin & H Garber (Eds.), *Research on exemplary schools* (pp. 3–30). Orlando, FL: Academic.
- Murnane, R, & Willett, JB (Eds.). (2011). *Methods matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press.
- Postlethwaite, TN, & Ross, KN. (1992). *Effective schools in reading. Implications for Educational Planners*. The Hague: IEA.
- Purkey, SC, & Smith, MS. (1983). Effective schools: A review. *The Elementary School Journal*, 83(4), 426–452.
- Raudenbush, S, Bryk, A, Cheong, YE, Congdon, R, & du Toit, M. (2004). *HLM6: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International, Inc.
- Rutkowski, L, & Rutkowski, D. (2009). Trends in TIMSS responses over time: Evidence of global forces in education? *Educational Research and Evaluation*, 15(2), 137–152.
- Sandoval-Hernandez, A. (2008). School effectiveness research: A review of criticisms and some proposals to address them. In *Education (Special Issue)* (pp. 31–44).
- Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, 1, 61–80.
- Scheerens, J. (2000a). *Improving school effectiveness*. Paris: UNESCO, International Institute for Educational Planning. (Fundamentals of Educational Planning Series No. 68).
- Scheerens, J. (2000b). *School effectiveness in developed and developing countries: A review of the research evidence*. World Bank. <http://www.worldbank.org/education/schools>.
- Scheerens, J. (2004). *Review of School and Instructional Effectiveness Research. Paper commissioned for the EFA Global Monitoring Report, 2005*. The Quality Imperative, UNESCO.
- Scheerens, J, & Bosker, RJ. (1997). *The foundations of educational effectiveness*. Oxford: Elsevier.
- Scheerens, J, Vermeulen, CJA, & Pelgrun, WJ. (1989). Generalizability of school and instructional effectiveness indicators across nations. In BPM Creemer & J Scheerens (Eds.), *Development in school effectiveness research. Special issue of the International Journal of Educational Research*. Oxford: Pergamon Press. 13(7), 789–800.
- Schmidt, WH, Raizen, SA, Britton, ED, Bianchi, LJ, & Wolfe, RG. (1997a). *Many visions, many aims. Cross-national investigation of curricula intentions in school science* (Vol. A). Dordrecht: Kluwer.
- Schmidt, WH, Raizen, SA, Britton, ED, Bianchi, LJ, & Wolfe, RG. (1997b). *Many visions, many aims. Cross-national investigation of curricula intentions in school science* (Vol. B). Dordrecht: Kluwer.
- Schmidt, WH, Wung, H, & Logan, LS. (2001). *Why schools matter: Cross national comparison of curricula and learning*. San Francisco, CA: Jossey Bass.
- Schwab, JJ. (1962). The concept of the structure of a discipline. *Education Record*, 43, 197–205.
- Seidel, T, & Shavelson, RJ. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
- Shulman, LS, & Tamir, P. (1973). Research on teaching in the natural sciences. In RWM Travers (Ed.), *Second handbook of research on teaching* (pp. 1098–1148). Chicago: Rand McNally.
- Stallings, J. (1985). Effective elementary classroom practices. In MJ Kyle (Ed.), *Reaching for excellence. An effective schools sourcebook*. Washington, DC: US Government Printing Office.

- Stevenson, HW, Stigler, JW, Lucker, GW, Lee, S, Hsu, CC, & Kitamura, S. (1987). Classroom behavior and achievement of Japanese, Chinese, and American children. In R Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 3, pp. 153–191). Hillsdale, NJ: Erlbaum.
- Stigler, JW, Gonzales, P, Kawanaka, T, Knoll, S, & Serrano, A. (1999). *The TIMSS videotape classroom study methods and findings from an explanatory research project on eighth-grade mathematics instruction in Germany, Japan and the United States (NCEC-99)*. Washington, DC: US Department of Education, National Center for Education Statistics.
- Tomlinson, CA, Brighton, C, Hertberg, H, Callahan, CM, Moon, TR, Brimijoin, K, Conover, LA, & Reynolds, T. (2003). Differentiation instruction in response to student readiness, interest and learning profile in academically diverse classrooms. A review of literature. *Journal for the Education of the Gifted*, 27, 119–125.
- Von Secker, CE, & Lissitz, RW. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching*, 36(10), 1110–1128.
- Von Secker, CE. (2002). Effect of inquiry-based teacher practices on science: Excellence and equity. *The Journal of Educational Research*, 15(3), 151–160.
- Walberg, HJ. (1984). Improving the productivity of American schools. *Educational Leadership*, 41, 19–27.
- Wang, MC, Haertel, GD, & Walberg, HJ. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63(3), 249–294.
- Wrigley, T. (2004). School effectiveness: The problem of reductionism. *British Educational Research Journal*, 30(2), 227–244.
- Zabulionis, A. (2001). Similarity of mathematics and science achievement of various nations. *Educational Policy Analysis Archives*, 9, 33.

doi:10.1186/2196-0739-1-2

**Cite this article as:** Zuzovsky: What works where? The relationship between instructional variables and schools' mean scores in mathematics and science in low-, medium-, and high-achieving countries. *Large-scale Assessments in Education* 2013 1:2.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---