Achievement Data in IEA Studies and Simpson's Paradox

Ruth Zuzovsky, Tel Aviv University and Kibbutzim College of Education Technology and the Arts, ruthz@post.tau.ac.il
David Steinberg, Tel Aviv University, dms@post.tau.ac.il
Zipi Libman, Kibbutzim College of Education Technology and the Arts, Zipi.libman@gmail.com

Abstract

This paper is meant to highlight the occurrence of Simpson's Paradox when using aggregated data obtained from two IEA studies in Israel, while ignoring the effect of a powerful intervening variable in the Israeli context – the ethnicity factor. It will demonstrate faulty conclusions on the absence of relationships between a contextual variable and achievement, while such relationships do exist or conclusions on the existence of such relationships – while in reality they do not exist.

Our intention in writing this paper is to draw the attention of our fellow researchers to similar faulty inferences they might come across when analyzing their local database within the scope of IEA studies.

Keywords: *Achievement data, Simpson's Paradox*

Introduction

International comparative studies of educational achievement such as the IEA Studies are concerned with measuring the outcomes of national education systems and with the relationship between these outcomes and the national context of learning in schools, classrooms and students' homes.

Using equivalent representative samples of the student population and valid measures of outcomes in main school subjects and assuring high alignment between the national curricula and the international assessment framework, these studies intend to rank countries according to their educational attainment and provide information on conditions and practices that are leading to superior results in some of them.

These two main uses of data obtained from international comparative educational studies, often referred to as "descriptive" and "explanatory" are meant to improve educational policy making. Unfortunately, due to the complexity of the school phenomenon and their cross-sectional design and

the aggregated nature of the data they provide, the international comparative studies cannot provide decisive evidence for choosing among alternative policy interventions, nor can they support "weak" or "synthetic" causal inferences (Smith, 2002) often drawn from detected relationships between contextual variables and student achievement as these are often based on aggregated country data.

Warnings about the usage of one number scores, typically the mean of a country for both descriptive and explanatory purposes, were raised by Raudenbush and Kim (2002). They point to the fact that it is possible that country A will have a higher mean score than country B even though every subgroup (gender, race, etc.) in country A, does worse than the corresponding subgroup in country B. It is also possible that the relationship between an independent contextual variable and achievement inferred from the aggregated data of the entire country will not fit or even reverse that inferred from disaggregated data of subgroups within this country.

Such surprising findings are caused by a well-known statistical paradox. This paradox, first described by Yule in 1903, is named after the person who later developed and popularized it – Simpson's Paradox (Blyth, 1972; Freedman, Pisani, & Purves, 1998; Rinott & Tam, 2003; Simpson, 1951; Wainer, 1986a,b; Wainer & Brown, 2004).

The paradox results from the existence of intervening covariates (a confounder, lurking variable) that, like the relevant contextual variables, is also associated with achievement. When subjects belonging to subgroups defined by the categories of the intervening covariate are unequally distributed at different levels of the relevant contextual variable, achievement results for the entire population are weighted incorrectly and the relationship between the contextual variable and achievement as inferred on the basis of the aggregated data can be misleading.

Wainer and Brown (2004) use findings from the National Assessment of Educational Progress of 8th grade students in Nebraska and New Jersey to demonstrate this paradox. As seen in the chart below, eighth grade students in Nebraska scored six points higher in mathematics than their counterparts in New Jersey. Yet, both Whites, Blacks and other students did better in New Jersey:

State		White	Black	Other
Nebraska	277	281	236	259
New Jersey	271	283	242	260
		Proportion of		Population
		White	Black	Other
Nebraska		87%	5%	8%
New Jersey		66%	15%	19%

This counter-intuitive, but empirical evidence, is an example of Simpson's Paradox. In this example, ethnicity, which is highly associated with achievement, is the intervening variable. As a much greater percentage of Nebraska's 8th grade students (87%) are from higher scoring White populations than in Jew Jersey (only 66%), their scores contribute more to the total.

Attention to the effect of Simpson's Paradox on inferences drawn from large scale surveys of educational achievement was given in the United States in several cases. Terwilliger and Schield (2004) reported, on the basis of NAEP data for 2000 and 2002, the occurrence of approximately 100 instances of Simpson's paradox due to three intervening variables: family income, school location and ethnicity. In these cases the difference between the mean scores of two states was found to have opposite signs to those found in each state's subgroups (as defined by the categories of the intervening variables). Other reports that highlight the effect of Simpson's Paradox are based on SAT data bases or other data in the United States (Bracy, 1997, 2006; Wainer, 1986; Wainer & Brown, 2004).

Not too many researchers have dealt with the effect of Simpson's Paradox in the context of IEA studies. As mentioned before, attention to it was given by Raudenbush & Kim, 2002, and by Rowan, 2002. These scholars recommended to be aware of the existence of confounding variables and take precautions that will help not to misinterpret aggregated mean differences.

The effect caused by Simpson's Paradox was noticed in Israel as well while analyzing PIRLS 2006 and TIMSS 2007 data. Looking for relationships between several contextual variables and learning outcomes, we realized that achievement gaps of students exposed to high levels of a contextual variable and of those exposed to low levels of a contextual variable in the entire sample did not fit those computed for ethnic subgroups in the country, and also differed a lot from those computed on the basis of the international averages. In seeking an explanation for this puzzling phenomenon, Simpson's Paradox is a natural choice.

In the Israeli reality, a powerful intervening variable is ethnicity. It distinguishes between two ethnic subgroups that differ a lot in their attainment. The majority and better performing students' study in Hebrew-speaking schools and those who study in Arabic-speaking schools have lower levels of attainment. This variable is also an indicator of other existing social, economic and cultural differences between the two ethnic groups that are all highly associated with achievement.

As the proportions of subjects belonging to each ethnic group at each level of the contextual variables were found to be extremely unequal, achievement results for the entire population were weighted incorrectly and the achievement gaps of students exposed to the distal categories of the contextual variable in the entire population were not compatible with those computed for the two ethnic groups.

Purpose and Method

This paper is meant to highlight the occurrence of Simpson's Paradox when using aggregated data obtained from two IEA studies in Israel, while ignoring the effect of a powerful intervening variable in the Israeli context – the ethnicity factor. It will demonstrate faulty conclusions on the absence of relationships between a contextual variable and achievement, while such relationships do exist or conclusions on the existence of such relationships – while in reality they do not exist.

Our intention in writing this paper is to draw the attention of our fellow researchers to similar faulty inferences they might come across when analyzing their local database within the scope of IEA studies.

We will report on only a few of the instances that we found when recently analyzing PIRLS-2006 and TIMSS-2007 data in Israel and describe the measures we took in order to overcome such problems. In general, these measures were:

1. Controlling for the effect of the intervening variable (using analysis of covariance.
2. Adjusting for the unbalanced distribution of subjects belonging to the categories of the intervening variable (ethnicity) within each level of the relevant contextual variable using a common demographic mixture of Hebrew-speaking and Arabic-speaking student in Israel – a technique known as "standardization" (Wainer & Brown, 2004). This new weighting enables to generate new achievement scores for each level of the relevant contextual variable that adjusts for the unbalanced distribution of subjects to the categories of the confounder.

In elaborating on the examples, we will refer to tables that provide the following data for each example: The distribution of the subjects at each level of the relevant contextual variable for the entire sample and for Hebrew-speaking students and Arabic-speaking students separately; Breakdown of the **observed weighted international** mean scores for students at each level of the relevant contextual variable and **observed weighted Israeli** mean scores for the entire sample and for subgroups followed by **adjusted weighted Israeli** mean scores for the entire sample using the two methods: adj1 – using the Ancova procedure and adj2 – using the standardization technique. Gaps between the adjusted mean score of students belonging to the distal levels of the contextual variable will be used to support inference on the relationship between these variables and achievement, and will be compared to inferences drawn from unadjusted scores.

Findings

Examples Based on PIRLS 2006 Data in Israel

Example 1: *The Relationship Between Early Home Literacy Activities and Reading Achievement of 4th Graders*

This variable is an index (EHLA) constructed from parents' responses on the frequency of being engaged with their child prior to entry into a primary school in reading books, telling stories, singing songs, playing with alphabet toys, playing word games, reading signs and labels aloud.

The average parents' responses on a scale of 1 (never) to 3 (often) were cut into three categories indicating High, Medium or Low frequency of early home literacy activities.

The distribution of subjects in each category of this variable in the entire sample and in each group defined by the intervening variable is shown in Table 1. This distribution is clearly different in

the two ethnic groups. As ethnicity is related to achievement, and achievement is much higher for Hebrew-speaking students, the basic conditions for Simpson's Paradox to occur are present.

Indeed, the observed achievement data for each ethnic group show a positive monotone relationship, however, for the entire population, this relationship is negative, i.e., students who are exposed to high levels of EHLA, achieve less than students exposed to low levels of EHLA.

Insert Table 1 about here

How can this be? What happens when we combine the two ethnic groups? As can be seen, the combined (entire sample) low group is almost all Hebrew-speakers, the combined medium group is 87% Hebrew-speakers and the combined **high** group is only 2/3 Hebrew-speakers. As a result, the average achievement for the combined high group is "brought down" by the large proportion of Arabic-speaking students. In contrast, the average achievement for the **low** group is almost equal to the average of the Hebrew-speaking students, the majority in this group. The result is a reverse of the positive monotone relationship that is seen in each ethnic group.

The adjustment made that takes into account the influence of ethnicity and the unequal proportions of subjects belonging to each of the ethnic groups at each level of the early home literacy activity generated new scores for the combined group and the achievement in the combined group of students under frequent and less frequent exposure to early home literacy activities shows gaps in favor of students exposed to frequent early home literacy activity. These gaps are more sensible and are similar to those found according to international data.

Example 2: *The Relationship Between Students Reading Aloud to their Class and Reading Achievement*

This example looks at a variable that plays a role at the class level. The data is obtained from students' responses to the question: *How often do you read aloud to students in your class?* on a scale ranging from 1 (every lesson or almost every lesson) to 4 (never or almost never). Table 2 presents data referring to this example.

Insert Table 2 about here

The key features of Simpson's Paradox appear here. Subjects in the two ethnic groups have different distribution with respect to the categories of the independent variable.

In the Hebrew-speaking population, there is almost no relationship between reading achievement and the frequency of reading aloud to the whole class. In the Arabic-speaking population, there is a monotone positive relationship between frequent reading aloud and achievement. As almost 55% of the students who read aloud to their class every day are less-achieving, Arabic-speaking students, their weight "brings down" the aggregated average score for the combined group at this frequency. The small proportion of the Arabic-speaking group who never or almost never read aloud to the class only slightly "brings down" the achievement of the combined

group at this frequency. As a result the observed gap for the entire sample is in favor of students who do not read aloud in class. However, after the two adjustments the gap for the entire sample is reversed showing higher attainment for those who read aloud in every lesson – a result which is much more reasonable than the one obtained using the aggregated observed scores before adjustment. Frequent reading aloud to the whole class is an important activity for both sectors and is more strongly associated with reading achievement for Arabic-speaking students.

Example 3: *Relationship Between Availability of School Resources and Student Reading Literacy Attainment*

The last example deals with a school level index that describes responses of principals to a question regarding the effect of shortage or inadequacy of a list of material and human resources in schools. This index has three categories – high level, i.e., high availability of school resources, medium level and low level. This variable is regarded as positively related to achievement and is often manipulated by policymakers. Most of the students studying in affluent schools (high on this index) are Hebrew-speaking (85% vs. 15% of the Arabic-speaking students) while in schools where availability of resources is low – 45% of students are Arabic-speaking and 55% are Hebrew-speaking students (see Table 3). A positive monotonous relationship between levels of this index and reading literacy attainment is observed in the aggregated data both according to international averages and the Israeli means. However, this relationship is not consistent in the two ethnic groups. Whereas in Hebrew-speaking schools, availability of school resources seems not to be related to reading achievement, in the Arabic-speaking schools it is positively related. These findings are at odds with those revealed from the aggregated data and might lead to a differential policy of funding schools in the two sectors. Adjusting the scores, taking into account the representation of subjects from the two sectors at each level of the availability of school resources index, changes the picture and reduces the impact of this variable on achievement in the entire sample.

Insert Table 3 about here

Examples Based on TIMSS-2007 Data in Israel

Example 1: *The Relationship Between Parents Origin and Mathematics and Science Achievement*

Student responses to a question about their parents' country of origin were recoded into three categories that described whether both parents of a student were born in the country (native), only one of them, or neither of them (immigrants). This variable is known to be related to student attainment. However, this relationship is also distorted due to the intervening "ethnicity" variable also associated with student attainment.

Insert Table 4 about here

When the distribution of subjects belonging to each ethnic group in each of the categories of the "born in country" variable is different, there is a possibility for Simpson's Paradox to occur. Two-thirds of students whose parents were born in Israel are Hebrew-speaking and one-third Arabic-speaking. Almost all students whose parents are immigrants are Hebrew-speaking students, so the mean achievement of the entire sample is similar to that of the Hebrew-speaking group alone. These patterns appear in both mathematics and in science.

The observed achievement gap between students with both parents born in Israel and those where neither parent was born in Israel, on the basis of aggregated data is -2 in mathematics and only one score point in science. Inferring from the observed scores for the entire population, it seems that in Israel, this variable is not related to students' achievement. However, when calculated for each ethnic group separately, this variable shows a positive achievement gap in mathematics of about 1/5 of a standard deviation in favor of students with native parents in both sectors, similar to the one found according to the international average score. In science this gap is 16 score points in favor of students with native parents for Hebrew-speaking students and 40 score points for Arabic-speaking students. These results contradict the conclusion drawn on the basis of the aggregated data.

The adjustments made using Ancova to control the effect of the intervening variable (ethnicity) and the adjustment made using the "standardization" method in which scores were calculated based upon a common ethnicity mixture (the Israeli overall proportion of Hebrew-speakers and Arabic-speakers that responded to this question (74.3% vs. 25.7%) yields more sensible results – both being similar to international findings and also within the range of the gaps observed in the two ethnic groups.

Using aggregated data in this case clearly leads to faulty inferences, especially if the question we are trying to answer is whether there exists a possible causal relationship between the origin of students' parents and mathematics and science achievement. Here adjustment for the effect of the intervening variable is essential to provide a valid answer.

Example 2: *The Relationship Between an Instructional Variable and Student Achievement*

An example of the relationship between a highly regarded quality type of mathematics instruction and student achievement will be used to illustrate the distortion due to Simpson's Paradox on inferences drawn from aggregated data at the class level.

This mode of instruction refers to the frequency reported for *students work on their own in solving mathematical problems*. Students responded on a scale of 1 (every or almost every lesson) to 4 (never or almost never). Table 5 presents data in regard to this example. Here, too, the distribution of subjects belonging to each ethnic group is different at each level of the instructional variable. The majority of students who *work on mathematical problems on their own every or almost every lesson* are Hebrew-speaking students (83% vs. 17%), while the majority of students *who never solve problems on their own* are Arabic-speaking students (60% vs. 40%).

Insert Table 5 about here

This unbalanced distribution of subject belonging to each ethnic group at the different levels of the instructional variable provides conditions suitable for Simpson's Paradox to occur.

A gap between the observed scores of students solving mathematical problems on their own, every or almost every lesson and those who never accomplish this, amounts, according to international averages, to 40 score points. In Israel, this gap in the entire population is 67 score points and it exceeds the gaps found in the two subgroups – 54 score points in Hebrew-speaking schools and only 24 score points in Arabic-speaking schools.

When adjusting the scores using the two techniques discussed above, the gap for the entire sample is reduced to 41 (Ancova) or 47 (standardization) score points and is now similar to the reported international average. This reduced gap describes more accurately the relationship between this mode of instruction and mathematics achievement for the entire sample.

Frequent opportunities given to students to solve mathematical problems on their own in both sectors are related to mathematics achievement, with a stronger correlation for Hebrew-speaking students. These differences might indicate that, in addition to Simpson's Paradox, there might be an interaction effect between this mode of instruction and student characteristics that distinguishes between the two subgroups.

Example 3: *The Relationship Between School Climate and Student's Achievement*

The last examples from TIMSS-2007 data are related to the relationship between a school level variable – a measure of a positive school climate and student achievement.

The school climate index is based on averaging principals' responses to eight questions about their school: teachers' job satisfaction, teachers' understanding of schools' curricular goals, success in implementing the school curriculum, teachers' expectations for student achievement, parental support, parental involvement in school activities, students regard for school property and their desire to do

well in school. These favorable attributes are regarded according to school effectiveness research as good indicators of school attainment. The index is dived into three categories: High, medium and low learning climate, with high indicating a positive climate.

Insert Table 6 about here

The distribution of students by ethnic group differs at each category of the school climate variable differs, so again Simpson's Paradox might appear. Most students in the high category of school climate are Hebrew-speaking students, versus only one third at the low level of this variable.

According to international data, students in schools characterized by high levels of school climate gain in mathematics – 39 score points more, on average, than students in schools with low levels of school climate. This gap is even larger in science where it reaches 45 score points.

In Israel, the gap between the observed scores of the distal categories of the school climate variable for the entire sample is 61 score points in mathematics and 52 score points in science, more than those detected by the international averages. Moreover, the aggregate gaps are beyond those revealed for each ethnic group. One might infer that in Israel the school climate variable is very "effective".

After controlling for the effect of the confounding variable, ethnicity, the adjusted scores computed through the two methods of adjustment yield smaller achievement gaps between the two distal categories of the school climate variable – about 26 score points in mathematics and about 22 (Ancova) 21 (standardization) score points in science and they are now similar to the observed international gaps, leading to a less enthusiastic belief in the power of this very popular school attribute.

Concluding Remarks

The ultimate utility of comparative educational achievement studies is to support the improvement of policy-making. Policy decisions are based on causal inferences, but unfortunately most comparative surveys of educational achievement are cross-sectional studies that cannot support such causal inferences. Only complicated longitudinal experimental trials that are very expensive might serve this purpose (Raudenbush & Kim, 2002).

Being aware of these limitations, this paper pays attention to another shortcoming in data obtained in international comparative studies of educational achievement that can lead to invalid causal inferences. This often occurs when using one-number aggregate means as a basis for inferences on causal relationships while ignoring intervening variables that stratify the aggregated data.

In writing this paper we respond to a call addressed by Raudenbush and Kim (2002) aimed at analysts working on international comparative studies amongst them, a call to study confounding variables that define subgroups in a population and take necessary precautions to control for their

effect on inferences drawn from aggregated data of the entire population. We use examples of the distortion in inferences that occurs due to ignoring powerful intervening variables of ethnicity and provide ways to adjust for those effects and remove the distortion. However, one must take into consideration that adjusting data for confounders is controversial.

In the United States, the National Assessment Governing Board (NAGB) warn against using adjusted or predicted scores based on ethnic and other demographic characteristics, or on opportunity to learn variables. The board claims that these adjustments convey a message of admitting an inequality. The Board notes that any adjusted or predicted scores could be subject to serious methodological and political challenges and would be contrary to the strong commitment to encouraging high standards for all children (NAGB, **199…).**

Wainer and Brown (2004) mention other reasons for not adjusting aggregated data. They claim that the reason for adjustment is dependent on the question we are trying to answer. If the question is related to aggregates such as: "I want to open a business. In which State will I find a higher proportion of high scoring mathematics students?", unadjusted mean achievement scores of the entire population will give us an appropriate answer for such a question.

On the other hand, the question of interest might be: "I want to enroll my children in school. In which State are they likely to do better in mathematics?" Here, attention to the race of a child – no matter which one – will give the relevant answers and it makes sense to adjust for ethnicity.

In the context of analyzing IEA studies in Israel, when the questions posed are related to the relationship between relevant contextual variables and achievement, there is a need to adjust for the powerful intervening covariate of ethnicity.

Acknowledging the role of international comparative studies of educational achievement in educational policy making, one should react both to the question in focus and take precautions against naïve inferences that can be drawn from aggregated data.

References

Blyth, C.R. (1972). On Simpson's Paradox and the sure-thing principle. *Journal of American Statistical Association, 67* (338), 364-366.

Bracy, G.W. (1997). *Setting the record straight: Responses to misconceptions about public education in the United States.* Alexandria, VA: Association for Supervision and Curricular Development.

Bracy, G.W. (2006). *Reading educational research: How to avoid getting statistically snookered.* Heinemann.

Freedman, D., Pisani, R., & Purves, R. (1998). Statistics (3rd ed.). W.W. Norton, ISBN 0-393-97083-3.

Raudenbush, S.W. & Kim, J. (2002). Statistical issues in analysis of International comparisons of educational achievement. In A.C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp.267-294). Board of International Comparative Studies in Education National Research Council, National Academy Press, Washington DC.

Rinott, Y. & Tam, M. (2003). Monotone regrouping, regression, and Simpson's Paradox. *The American Statistician, 57,* 139-141.

Rowan, B. (2002). Large scale cross-national surveys of educational achievement. Promises, pitfalls, and possibilities. In A.C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 321-349). Board of International Comparative Studies in Education National Research Council, National Academy Press, Washington DC.

Smith, M. (2002). Drawing inferences for national policy from large-scale cross-national surveys. In A.C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 295-312). Board of International Comparative Studies in Education National Research Council, National Academy Press, Washington DC.

Terwilliger, S., & Schield, M. (2004). *Frequency of Simpson's Paradox in NAEP data.* Presented at AERA, September, 2004.

Wainer, H. (1986). Minority contributions to the SAT score turnaround: An example of Simpson's Paradox. *Journal of Educational Statistics, 11,* 239-244.

Wainer, H., & Brown, L. (2004). Two statistical paradoxes in the integration of group differences: Illustrated with medical school admission and licensing data. *The American Statistician, 58,* 117-123.

Yule, G.U. (1903). Notes on the theory of association of attributes of statistics. *Biometrika, 2,* 121-124.

Table 1: PIRLS – Frequency Distribution of Subject at Each Level of Early Home Literacy Activities Variable for the Entire Sample and for Each Ethnic Group and Breakdown of Observed and Adjusted Reading Literacy Mean Scores

Early Home Literacy Activities (ehla2001) Frequency							Scores					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Population		Entire Population	Intnl
	n	%	*n*	%	*n*	%	Ob	Ob	Ob	Adj1	Adj2	Ob
High	1759	72.7	1131	64.3	627	35.7	569	449	526	547	534	515
Medium	539	22.3	469	87.0	70	13.0	549	412	531	525	509	494
Low	122	5.0	115	94.4	7	5.7	539	406	531	516	500	475
All	2420	100	1715	70.9	704	29.1						
Δ							30	43	-5	31	34	40

Table 2: PIRLS – Frequency Distribution of Subjects at Each Level of Reading Aloud in the Class Variable for the Entire |Sample and for each Ethnic Group and Breakdown of Observed and Adjusted Reading Literacy Mean Scores

Student Reads Aloud to Class (hc2) Frequency							Scores					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Population		Entire Population	Intnl
	n	%	*n*	%	*n*	%	Ob	Ob	Ob	Adj1	Adj2	Ob
Every/almost every day	1038	27.9	460	44.3	578	55.7	547	443	490	526	518	486
Once/twice a week	1335	35.9	1009	75.6	326	24.4	549	427	519	518	515	506
Once/twice a month	783	21.0	699	89.3	84	60.7	559	410	543	524	517	510
Never/almost never	565	15.2	499	88.3	66	11.7	543	395	526	509	501	494
All	3721	100	2667	71.7	1054	28.3						
Δ							4	48	-36	17	17	-10

Table 3: PIRLS – Frequency Distribution of Subjects at Each Level of the Availability of School Resources Index (ASR) for the Entire Sample and for Each Ethnic Group and Breakdown of Observed and Adjusted Reading Literacy Mean Scores

Availability of School Resources (ASR)							Scores					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Sample		Entire	Intnl
	n	%	*n*	%	*n*	%	Ob	Ob	Ob	Adj1	Adj2	Ob
High	1262	36.7	1078	85.4	184	14.6	546	445	532	513	517	505
Medium	1390	40.5	951	68.4	439	31.6	545	422	507	508	511	496
Low	784	22.8	433	55.2	351	44.8	549	426	494	511	514	476
All	3436	100	2462	71.7	974	28.3						
Δ							-3	19	38	2	2	29

Table 4: TIMSS – Frequency Distribution of Subjects at Each Level of Parents Born in Country Variable and Breakdown of Observed and Adjusted Achievement Gaps in Mathematics and Science Mean Scores

Parents Born in Country (BSD6BORN) Frequency							Mathematics Score						Science Score					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Populaton		Entire Pop.	Intnl.	Hebrew-Speaking	Arabic-Speaking	Entire Population		Entire Pop.	Intnl.
	n	%	*n*	%	*n*	%	Ob	Ob	Ob	Adj1	Adj2	Ob	Ob	Ob	Ob	Adj1	Adj2	Ob
Both	1973	63.1	1289	65.3	684	34.7	491	422	467	479	473	454	492	438	473	483	478	470
Only one	510	16.3	421	82.7	88	17.3	493	374	472	472	462	439	493	375	473	472	463	453
Neither	644	20.6	612	95.0	32	5.0	472	400	469	458	454	430	476	398	472	463	456	442
All	3126	100	2322	74.3	804	25.7												
Δ							19	22	-2	21	19	24	16	40	1	20	22	28

(1) Adjusted scores using Ancova
(2) Adjusted scores through standardization

Table 5: TIMSS – Frequency Distribution of Subjects at Each Level of Students Work Out Problems on their Own Variable and Breakdown of Observed and Adjusted Achievement Gaps in Mathematics Scores

Students Work Out Problems on their Own (BS4MHWPO) Frequency							Mathematics Score					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Sample		Entire	Intnl.
	n	%	*n*	%	*n*	%	Ob	Ob	Ob	Adj1	Adj2	Ob
Every, almost every lesson	1762	56.6	1465	83.1	297	16.9	499	427	487	474	481	461
Half lesson	841	27.0	599	71.2	242	28.8	472	412	454	449	456	456
Some lessons	385	12.3	200	51.9	185	48.1	449	407	429	435	438	442
Never	125	4.0	50	39.7	76	60.3	445	403	420	433	434	421
All	3114	100	2314	74.3	800	25.7						
Δ							54	24	67	41	47	40

(1) Adjusted scores using Ancova
(2) Adjusted scores through standardization

Table 6: TIMSS – Table 5: TIMSS – Frequency Distribution of Subjects at Each Level of School Climate Variable and Breakdown of Observed and Adjusted Achievement Gaps in Mathematics and Science Mean Scores

School Climate (PPSC) Frequency							Mathematics Scores						Science Scores					
	Entire Sample		Hebrew-Speaking		Arabic-Speaking		Hebrew-Speaking	Arabic-Speaking	Entire Sample		Entire	Intnl.	Hebrew-Speaking	Arabic-Speaking	Entire Sample		Entire	Intnl.
	n	%	*n*	Ob	Ob	%	Ob	Ob	Ob	Adj1	Adj2	Ob	Ob	Ob	Ob	Adj1	Adj2	Ob
High	794	26.3	664	81.2	149	18.8	497	445	488	474	483	149	499	469	493	481	490	498
Medium	1997	66.2	1441	72.2	556	27.8	485	403	462	455	462	465	485	414	465	459	465	462
Low	224	7.4	79	35.3	145	64.7	479	399	427	448	457	445	490	415	441	459	469	427
All	3014	100	2164	71.8	850	28.2												
Δ							18	46	61	26	26	39	9	13	52	22	21	45

(1) Adjusted scores using Ancova
(2) Adjusted scores through standardization