Performance in Statistics : An International Perspective on Gender Differences

Gila Hanna - Toronto, Canada

1. Background : The Second International Mathematics Study

The purpose of this paper is to investigate gender differences in achievement in statistics, making use of the Population A data of the Second International Mathematics Study (SIMS) conducted by the International Association for the Evaluation of Educational Achievement (IEA). In IEA terms, Population A means all the students in the grade in which most students attain the age of 13.0 to 13.11 years by the middle of the school year.

The SIMS study used a very large random sample stratified by region and school type in each of the 20 participating "countries", which were in effect 20 different educational systems rather than countries. For example, Canada is represented by two educational systems, British Columbia and Ontario; Belgium by two systems, one Flemish and the other French.

A test consisting of 180 multiple-choice items covering five content topics (Arithmetic, Algebra, Geometry, Statistics and Measurement) was administered to the students. Out of these 180 items, 18 were designed to measure achievement in statistics. National centres were asked to indicate the importance accorded within their own country curriculum to each of the nine categories of the statistics test. Their responses indicated that the first four categories were judged to be important (collection, organisation, representation and interpretation of data), while the remaining five were judged less important (combinatorics, sample spaces, independent, mutually exclusive and complementary events).

The 18 items on the statistics test were distributed as follows:

3 items:	organisation of data	1 item:	probability
7 items:	representation of data	1 item:	unclassified
6 items:	mean, median, mode		

The IEA survey sought detailed information from each of the participating countries on two interrelated aspects of mathematics teaching: the intended curriculum and the implemented curriculum.

(i) The *intended* curriculum is that reflected in curriculum guides, course outlines, syllabi and textbooks adopted by the educational system. An index of zero for the intended coverage in the case of statistics indicates that none of the statistics categories were intended to be included in the curriculum, while an index of 10 indicates that all the categories were indeed part of the curriculum (100% coverage).
(ii) The *implemented* curriculum is that actually taught in the classroom. Clearly, the teachers' selection of topics or patterns of emphasis may not be consistent with the intended curriculum. To identify the implemented curriculum, a questionnaire was developed for completion by the individual classroom teacher, eliciting highly specific information. The indices shown in Table 1 for the implemented curriculum are based on responses to this opportunity-to-learn questionnaire. Teachers were asked to rate each item on the test as to its appropriateness in light of the curriculum in the school, that is, whether the teachers thought that the content had been taught. No country reported a 100% or a 0% implementation. The data yielded a mean coverage of 58% with a range of 32% to 87% (Travers and Westbury, 1989, p.135). (A percentage between 30 and 40 is represented in Table 1 by an index of 3; country codes are explained in Table 2.)

As shown in Table 1, the reported indices for the implemented curriculum often differ from those reported for the intended curriculum. For example, Israel reported an index of 0 for the intended curriculum but 5 (52%) for the implemented curriculum.

TABLE 1
Indices of intended and implemented curriculum

Index	Country	
	Intended Curriculum	Implemented Curriculum
10	NZE SCO SWE USA	
9	CBC	
8	ENW FIN HKO HUN CON SWA	HUN SWA
7	JPN THA	USA JPN
6	NTH	CON ENW NZE NIG
5		FIN FRA ISR THA
4		CBC SWE
3		BFL LUX NTH
2	LUX FRA	
1	BFL BFR	
0	ISR	

Note: Data was not available for all the 20 countries. For the country codes see Table 2.

2. Performance by country and gender

Table 2 reports the number of girls and boys in each country and the mean and standard deviation of the percent of correct responses (p-c values) to the 18-item statistics subtest. The mean percent of correct responses for the test is obtained by averaging the percentage of correct answers calculated for each of the 18 items on the test.

TABLE 2
Descriptive data on the statistics subtest

	Girls			Boys		
	N	Mean	(sd)	N	Mean	(sd)
BFL : Belgium Flemish	718	64	(16)	652	61	(14)
BFR : Belgium French	874	57	(19)	1001	54	(18)
CBC : British Columbia	1080	63	(18)	1066	64	(17)
CON : Ontario	2401	58	(17)	2421	58	(17)
ENW : England	1338	60	(20)	1247	61	(19)
FIN : Finland	2091	59	(19)	2303	54	(16)
FRA : France	4629	57	(20)	3586	61	(19)
HKO : Hong Kong	2726	57	(16)	2822	57	(14)
HUN : Hungary	908	61	(22)	844	61	(21)
ISR : Israel	1651	52	(22)	1711	57	(20)
JPN : Japan	3924	72	(17)	4167	71	(16)
LUX : Luxembourg	1016	37	(19)	989	40	(18)
NTH : The Netherlands	2667	62	(19)	2769	65	(17)
NZE : New Zealand	2574	52	(18)	2629	52	(17)
NIG : Nigeria	389	34	(17)	1040	37	(18)
SCO : Scotland	627	60	(19)	729	60	(18)
SWA : Swaziland	485	37	(21)	414	35	(23)
SWE : Sweden	1660	56	(18)	1830	54	(15)
THA : Thailand	1833	49	(16)	1988	48	(16)
USA : USA	3452	54	(17)	3202	54	(16)
Total	37043	55	(20)	37410	55	(19)

The mean p-c values for country (boys and girls taken together) ranged from 35.5 to 71.5. A multivariate analysis of variance for the data of the 20 countries indicated that the country differences were statistically significant ($F_{(19,17)} = 25.62$, $p < 0.01$), but that the gender differences were not ($F_{(1,17)} = 0.15$); however there was a statistically significant interaction country by gender ($F_{(19,17)} = 6.65$, $p < 0.01$).

The scores varied from 50 and below (four countries, Luxembourg, Nigeria, Swaziland, and Thailand) to a high of over 70 (Japan). When the countries were ranked

according to their performance and grouped into low, medium and high scoring, analyses of variance for these three separate groups indicated no gender differences and no interactions country by gender within each of the groups of 4 and 15.

Clearly, the country results are very difficult to interpret, even after taking into account whether the mathematical content of the items was part of the intended curriculum or had actually been taught. Though Swaziland had an index of 8 (80%) for the implemented curriculum, the mean on the test was only 36%. On the other hand, Belgium (Flemish) reported an index as low as 3 (30%) for the implemented curriculum, but had a mean as high as 63%. The indices for the intended and implemented curriculum are thus limited in their explanatory power. Other factors, not measured by SIMS, also have a bearing on student achievement (Hanna, 1989). Under these circumstances comparisons across countries may not be meaningful. But since within each country the two sexes were taught the same curriculum, and should be equal on a number of relevant variables, such as socio-economic status, linguistic ability, and exposure to teaching methods, valid comparisons of the performance of boys and girls within each country can be made.

For each country the mean percentage of correct responses for girls was compared to that for boys, using the paired t-test with the item as the unit of analysis. As shown in Table 3, the results indicated that most differences did not reach statistical significance at the .005 level. The only differences that did reach significance were +5 percentage points in Finland (in the girls' favour), and -4, -5, and -3, in France, Israel and The Netherlands respectively (in the boys' favour).

TABLE 3

Mean percent differences of correct responses reaching statistical significance at the .005 level, by country

Country	Difference	Country	Difference
BFL : Belgium Flemish	--	JPN : Japan	--
BFR : Belgium French	--	LUX : Luxembourg	--
CBC : British Columbia	--	NTH : The Netherlands	-3
CON : Ontario	--	NZE : New Zealand	--
ENW : England	--	NIG : Nigeria	--
FIN : Finland	+5	SCO : Scotland	--
FRA : France	-4	SWA : Swaziland	--
HKO : Hong Kong	--	SWE : Sweden	--
HUN : Hungary	--	THA : Thailand	--
ISR : Israel	-5	USA : USA	--

Note: A positive difference in mean percent correct represents a higher mean percent for girls, and a negative difference a higher percent for boys; a dash (--) indicates that the difference did not reach statistical significance.

Since there were no gender differences in mean p-c values for the entire test, it seems interesting to explore the performance of boys and girls on individual items.

The differences between girls and boys for each of the 18 items (using the 15 countries in the mid-scoring range) are shown in Table 4. There were five items on which boys did better, one item with no differences, and 12 items on which girls did better.

TABLE 4
Differences between the genders in percent of correct responses
(18 items, data from 15 countries)

	Difference (%)	Number of Items
Boys > Girls	-3	1
	-2	1
	-1	3
No difference	0	1
Girls > Boys	+1	2
	+2	2
	+3	4
	+5	3
	+7	1

The item on which girls outperformed boys by 7 percentage points was:

Joe had three test scores of 78, 76, and 74, while Mary had scores of 72, 82, and 74. How did Joe's average compare with Mary's?

A. Joe's was 1 point higher
B. Joe's was 1 point lower
C. Both averages were the same
D. Joe's was 2 points higher
E. Joe's was 2 points lower

A close examination of the items on which girls did better does not seem to reveal any particular pattern. A number of items required students to organise data into frequency tables or find the average of a set of numbers; on some of these girls did better, while on seemingly similar items boys did better. In any case, the range of the differences is rather narrow and may not be of any educational significance.

Finally, the data were culled for information on the performance of New Zealand students. Though there were no overall mean differences between boys and girls in New Zealand, there were, as shown in Table 5, achievement differences across the items. The range of p-c values was from 32% (a difficult item) to 89% (an easy item).

TABLE 5
Percent of correct responses for New Zealand data by item and by gender

Item	Girls (%)	Boys (%)	Item	Girls (%)	Boys (%)
1	62	63	10	49	51
2	54	52	11	39	42
3	41	44	12	31	29
4	38	39	13	58	56
5	83	83	14	33	36
6	89	86	15	40	42
7	62	67	16	82	82
8	49	48	17	57	53
9	33	32	18	36	37

The two most difficult items were the following:

The arithmetic mean (average) of 1.50, 2.40, 3.75 is equal to:

A. 2.40 B. 2.55 C. 3.75 D. 7.65 E. None of these

There are 7,000,000 girls under the age of 21 in the country with a total population of 36,000,000. If a circle graph were drawn showing the distribution of the population, the angle in the sector representing girls under the age of 21 would have measure:

A. 7^{o} B. 20^{o} C. 21^{o} D. 70^{o} E. 72^{o}

New Zealand's mean percentage on the first item was 35, and on the second 32, while the mean percentages for 15 countries were 38 and 23 respectively. The percent of correct responses ranged from 21 to 71 for the first item, and from 11 to 40 for the second (for all 20 countries).

The two easiest items were one that showed a pie chart and asked the students to interpret it, and another that showed a bar graph and asked for information about it. New Zealand's mean percentages on these questions were 85 and 83 compared to 75 and 73 over the five countries, while for the 15 countries these were 75 and 73, and over all 20 countries scores ranged from 56 to 93, and from 57 to 88 respectively.

5. Conclusion

The purpose of this study was to determine if, among thirteen-year-olds, there are important gender differences in achievement in statistics. The results suggest that there are no such differences. It appears that girls and boys assimilate the subject matter taught in class equally well. Thus, it seems reasonable to conclude that gender is not a factor in the learning of statistics among thirteen-year-olds.

References

Hanna, G (1989) Mathematics achievement of girls in grade eight : results from twenty countries. *Educational Studies in Mathematics* **20**, 225-232.

Travers, K J and Westbury, I (eds) (1989) *The IEA Study of Mathematics I : Analysis of Mathematics Curricula.* Pergamon Press, Toronto.