Proceedings of the 8^{th} International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010. Vol. 2. pp. 17–27.

A Comparison of Informatics Skills by Genders of Hungarian Grammar School Students

Gábor Kiss

Óbuda University Department of Mechatronics and Cartechnics Engineering
Budapest, Hungary
e-mail: kiss.gabor@bgk.uni-obuda.hu

Abstract

An analysis of informatics skills by gender in secondary grammar schools was made with the help of a web based Informatics Test. In accordance with widespread opinions the hypothesis was that boys would better than girls. After the evaluation of the test results the correctness of this presumption has emerged. First the Kolmogorov-Smirnov-test was used to see if the groups showed standard normal distribution in answering the questions. The means of the correct answers by gender were examined using a Z-test with two parameters and the deviation quotient was to calculated revealing how much gender influences the difference of means. Significance level was 5% through the analysis. Significant divergence by gender was found regarding theoretical knowledge and programming showing the hypothesis correct in this two area.

Keywords: Informatics skills by gender, secondary grammar school, Hungary

1. Introduction

I made the comparison of the informatics skills by gender in a secondary grammar school, because I wanted to see if the hypothesis that boys are better than girls in informatics is true or not. Experiences show in higher education that boys are learning programming easier than girls. The audit was extended to other subjects too. I composed the test on the base of the National Curriculum and I analysed how effectively can students of different grades answer questions from different subjects [1]. 65 registered teachers participated in the test, all of them from different schools. According to the database of the test, the test was filled in by 872 students throughout Hungary (see Figure. 1, prepared using google-analytics) during school classes of the 9^{th}-12^{th} grades.The advantage of a web based Informatics Test is, that students can fill in it both at school or at home [2].

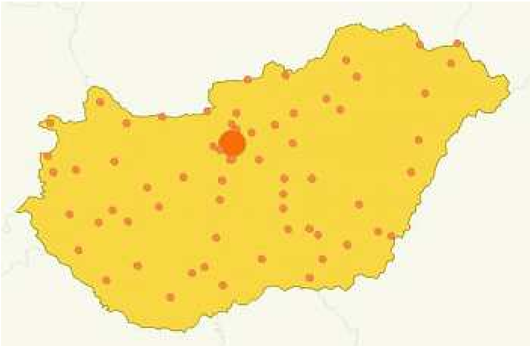

Figure 1: The distribution of schools using the Test

2. Test results

2.1. Number of the Participiants

The web based test was filled out by students of various grades summarized in the following table (Table 1.).

	Gender	
Grade	Boys	Girls
9	206	239
10	138	88
11	71	45
12	56	29

Table 1: The distribution of the students by gender and grade

2.2. Results by subjects

The following tables show the results by gender depending on grade and subjects (Table 2, Table 3, Table 4, Table 5).

		Boys			Girls		
Subject	Number of questions	Mean	%	Std. dev	Mean	%	Std. dev.
Theoretical knowledge	46	12,63	27,4%	6,27	10,31	22,4%	6,81
Word processing	14	5,84	41,7%	2,68	5,59	39,9%	2,81
Spreadsheet calculation	9	2,05	22,8%	1,68	2,11	23,5%	1,86
Database management	18	0,82	4,6%	1,92	0,64	3,5%	1,70
SQL	14	0,02	0,2%	0,21	0,03	0,2%	0,35
Programming	25	0,14	0,5%	1,01	0,31	1,2%	1,30

Table 2: Results by gender in 9^{th} grade

		Boys			Girls		
Subject	Number of questions	Mean	%	Std. dev	Mean	%	Std. dev.
Theoretical knowledge	46	12,38	26,9%	6,17	11,07	24,1%	7,03
Word processing	14	4,91	35,1%	2,76	5,70	40,7%	2,78
Spreadsheet calculation	9	2,57	28,6%	1,70	2,76	30,7%	1,98
Database management	18	1,61	8,9%	2,65	1,53	8,5%	2,40
SQL	14	0,22	1,6%	1,07	0,19	1,4%	0,74
Programming	26	0,67	2,6%	2,31	0,95	3,7%	3,19

Table 3: Results by gender in 10^{th} grade

		Boys			Girls		
Subject	Number of questions	Mean	%	Std. dev	Mean	%	Std. dev.
Theoretical knowledge	46	15,17	33,0%	8,56	10,71	23,3%	5,14
Word processing	14	6,14	43,9%	3,52	6,20	44,3%	2,98
Spreadsheet calculation	9	3,55	39,4%	2,33	3,02	33,6%	1,74
Database management	18	3,34	18,5%	3,18	3,16	17,5%	2,41
SQL	14	0,70	5,0%	2,41	0,00	0,0%	0,00
Programming	26	5,08	19,6%	5,42	0,98	3,8%	2,68
OOP	8	0,11	1,4%	0,55	0,00	0,0%	0,00

Table 4: Results by gender in 11^{th} grade

		Boys			Girls		
Subject	Number of questions	Mean	%	Std. dev	Mean	%	Std. dev.
Theoretical knowledge	46	15,61	33,9%	8,34	11,90	25,9%	5,85
Word processing	14	6,39	45,7%	2,95	6,10	43,6%	2,37
Spreadsheet calculation	9	3,32	36,9%	2,32	2,86	31,8%	1,64
Database management	18	2,68	14,9%	3,34	2,28	12,6%	2,42
SQL	14	0,54	3,8%	1,81	0,00	0,0%	0,00
Programming	26	2,00	7,7%	4,49	0,31	1,2%	1,67
OOP	8	0,16	2,0%	0,63	0,10	1,3%	0,41

Table 5: Results by gender in 12^{th} grade

It is apparent that results regarding the theoretical knowledge are found consequently between 20% and 30%. Boys achieved better means than girls, but the standard deviation is the highest among them.

When looking at word processing we find higher means at both genders (35-45%), and the values of spreadsheet calculation are the same as those of theoretical knowledge.

Database management gives better values in 10^{th} grade (˜18%), though according to the National Curriculum it should be taught earlier, but it run in this grade later. Results show that teachers has not enough time to teach SQL.

The results of subject programming shows some changes just in 11^{th} grade, but only among boys (˜19%). The 3,8% mean of girls is not significant. But in 12^{th} grade there is a significant fall of 7,7% among boys. Hereby we could draw the conclusion subject programming results show significant difference by gender, but first we have to make more analysis to verify it.

2.3. How many students had not learned the subject

We have seen earlier how could the students anwer the questions in the test. In the following we will see, how many students marked the subject as "I have not learned". By filling in the test opportunity is given chance to mark some subject as "I have not learned", in this situation the system will save this question with this answer and not pose further questions. The next tables show by grades and gender how many students have not learned the subjects.

	Boys		Girls	
Subject	Pers.	%	Pers.	%
Spreadsheet calculation	46	22,3%	51	21,3%
Database management	162	78,6%	200	83,7%
SQL	203	98,5%	237	99,2%
Programming	199	96,6%	223	93,3%

Table 6: How many students have not learned the subject in 9^{th} grade

	Boys		Girls	
Subject	Pers.	%	Pers.	%
Spreadsheet calculation	12	8,7%	10	11,4%
Database management	88	63,8%	56	63,6%
SQL	131	94,9%	82	93,2%
Programming	123	89,1%	77	87,5%

Table 7: How many students have not learned the subject in 10^{th} grade

Spreadsheet calculation is unknown by 1/5 of the students for although it should be taught already in 8^{th} grade according to the National Curriculum (Table 6). The database management is curricular in 9^{th} grade, but (look at table 9) ˜80% of the students have never heard about it, and this value for SQL is 100%. Programming shows wronger values than database management: more than 90% of the students have never heard about this subject which should begin already in the 7^{th} grade according to the National Curriculum.

The situation is better in 10^{th} grade: only have never learned spreadsheet calculation For database management the values are better too, "only" ˜63% of the students marked this subject as unknown. Changes in SQL and programming are not significant (Table 7.).

We see a significant difference by gender in results regarding programming in 11^{th} grade where 80% of the girls and 36,6% of the boys marked this subject as unknown (Table 8). This difference by gender is not so significant in 12^{th} grade (Table 9.).

	Boys		Girls	
Subject	Pers.	%	Pers.	%
Spreadsheet calculation	8	11,3%	4	8,9%
Database management	24	33,8%	9	20,0%
SQL	61	85,9%	45	100,0%
Programming	26	36,6%	36	80,0%
OOP	67	94,4%	45	100,0%

Table 8: How many students have not learned the subject in 11^{th} grade

	Boys		Girls	
Subject	Pers.	%	Pers.	%
Spreadsheet calculation	6	10,7%	2	6,9%
Database management	19	33,9%	11	37,9%
SQL	49	87,5%	29	100,0%
Programming	42	75,0%	28	96,6%
OOP	52	92,9%	27	93,1%

Table 9: How many students have not learned the subject in 12^{th} grade

2.4. Monitoring the standard normal distribution

To compare the means by gender we needed to confirm that the groups show standard normal distribution in answering the questions. We used the Kolmogorov-Smirnov test to decide [3]. As everyone answered the theoretical questions we used those as basic datas, because other subjects were markable as "I have never learned that".

The null hypothesis was that there was not significant difference between the standard normal distribution and the results of the genders. The monitoring was held on the α=5% significancy level in the two genders. The following table shows the maximum values of level α by genders (D_{max}), and the critical values of the Kolmogorov-test (D_{crit}). If the turnout is less than the critical value of the Kolmogorov-test, we may keep the null hypothesis, and the samples follow the standard normal distribution (Table 10.)

Grade	Gender	D_{max}	D_{crit}	Decision
9	Boys	0,04	0,09	keep the null hypothesis
9	Girls	0,09	0,12	keep the null hypothesis
10	Boys	0,07	0,12	keep the null hypothesis
10	Girls	0,1	0,14	keep the null hypothesis
11	Boys	0,12	0,16	keep the null hypothesis
11	Girls	0,09	0,19	keep the null hypothesis
12	Boys	0,08	0,09	keep the null hypothesis
12	Girls	0,11	0,24	keep the null hypothesis

Table 10: The results of Kolmogorov-Test

According to the table results by gender and grades show the calculated values less than the critical values of the Kolmogorov-test on the 5% significancy level, so the muster follows normal distribution.

2.5. Analysis of the means by subject

To get know if the means by subject are equal we need to make a Z-test [4].

The null hypothesis is that no significant difference exsist between the means by gender. The monitoring was held on the α =5% significancy level. The critical value of Z-test is between -1,96 and 1,96 on the α=5% significancy level. If the calculated value of Z-test is in this range, we keep the null hypothesis. The following tables show by grades and subject the calculated values of Z-test and the decision on keeping or not the null hypothesis.

Subject	Value of Z-test	Decision
Theoretical knowledge	-3,99	The means are not equal
Word processing	-15,60	The means are not equal
Spreadsheet calculation	-10,55	The means are not equal
Database management	-2,85	The means are not equal
SQL	-1,07	The means are equal
Programming	-2,66	The means are not equal

Table 11: Scores of the Z-test between genders in 9^{th} grade

We can say by α=5% significancy level that the only subject where boys and girls were on the same knowledge level was SQL in the 9^{th} grade (Table 11), which really means same ignorance level in this case. In other topics there was a significant difference between genders knowledge level.

Subject	Value of Z-test	Decision
Theoretical knowledge	0,28	The means are equal
Word processing	-10,16	The means are not equal
Spreadsheet calculation	-8,22	The means are not equal
Database management	-2,89	The means are not equal
SQL	-1,38	The means are equal
Programming	-1,75	The means are equal

Table 12: Scores of the Z-test between genders in 10^{th} grade

In 10^{th} grade we see the same knowledge level by gender in theoretical knowledge and in programming, similarly to SQL (Table 12).

Subject	Value of Z-test	Decision
Theoretical knowledge	6,75	The means are not equal
Word processing	-4,02	The means are not equal
Spreadsheet calculation	-4,42	The means are not equal
Database management	-2,71	The means are not equal
SQL	0,70	The means are equal
Programming	3,79	The means are not equal
OOP	0,11	The means are equal

Table 13: Scores of the Z-test between genders in 11^{th} grade

In 11^{th} grade we can say by α=5% significancy level that genders are on the same knowledge level in object oriented programming as in SQL (Table 13). This means students have never learned this subjects. We can see differences in knowledge level looking at other subjects (Table 13.).

Subject	Value of Z-test	Decision
Theoretical knowledge	7,96	The means are not equal
Word processing	-3,95	The means are not equal
Spreadsheet calculation	-3,26	The means are not equal
Database management	-0,92	The means are equal
SQL	0,54	The means are equal
Programming	1,54	The means are equal
OOP	-0,75	The means are equal

Table 14: Scores of the Z-test between genders in 12^{th} grade

In 12^{th} grade there are four subjects, based on the calcuated values, in which boys and girls are on the same knowledge level (Table 14.).

2.6. Measures of association

We have seen subjects and grades where there is significant difference between the knowledge of genders. Earlier we made sure standard normal distribution of the data exists in the case of both genders. So we can establish the influence of the chosen gender on the calculated means with the calculation of the Eta-squared (H^2) [5]. According to the previous tables there is no significant difference between the means got. In order to calculate the deviation quotient first we have to calculate the main mean ($\bar{\bar{x}}$).

We have to calculate the values of the variance between groups (S_k) and the variance within groups (S_b) where the standard deviation of musters appear in the formula.

The complete deviation quadrat (S) is the sum of the variance between groups and the variance within groups.

The following table shows the calculated values by subjects.

The value of the Eta-squared (H^2) is the quotient of the variance between groups and the complete deviation quadrat.

The calculated value shows in percentage, how the grouping by genders influences the difference between means. Calculating the square root of the Eta-squared is giving values between 0 and 1 (H), showing how strong the connection between the grouping by genders and the result achieved is. The higher the achievement, the stronger the connection. In the next table you can see the calculated values by subjects and the strength of the connection.

Subject	$\bar{\bar{x}}$	S_k	S_b	S	H^2	H	Connection
Theoretical knowledge	11,38	1,33	6,57	7,90	16,82%	0,41	Middling weak
Word processing	5,71	0,02	2,76	2,78	0,56%	0,07	No connection
Spreadsheet calculation	2,09	0,00	1,78	1,78	0,05%	0,02	No connection
Database management	0,72	0,01	1,81	1,82	0,46%	0,07	No connection
SQL	0,03	0,00	0,29	0,29	0,00%	0,00	No connection
Programming	0,23	0,01	1,17	1,18	0,61%	0,08	No connection

Table 15: How strong is the connection is between the grouping by genders in 9^{th} grade

According to the results the theoretical knowledge is the only value which is higher than 16% in 9^{th} grade, which means just middling weak connection between the chosen gender and this subject (Table 15.).

Subject	$\bar{\bar{x}}$	S_k	S_b	S	H^2	H	Connection
Theoretical knowledge	11,87	0,41	6,52	6,93	5,88%	0,24	Weak
Word processing	5,22	0,15	2,79	2,94	5,07%	0,23	Weak
Spreadsheet calculation	2,65	0,01	1,83	1,84	0,46%	0,07	No connection
Database management	1,58	0,00	2,57	2,57	0,05%	0,02	No connection
SQL	0,21	0,00	0,96	0,96	0,01%	0,01	No connection
Programming	0,78	0,02	2,67	2,69	0,70%	0,08	No connection

Table 16: How strong is the connection is between the grouping by genders in 10^{th} grade

We can see a connection between the chosen gender and Theoretical knowledge and word processing in 10^{th} grade, but it is very weak (˜5%) (Table 16.).

Subject	$\bar{\bar{x}}$	S_k	S_b	S	H^2	H	Connection
Theoretical knowledge	13,44	4,72	7,27	11,99	39,37%	0,63	Middling strong
Word processing	6,16	0,00	3,35	3,35	0,02%	0,02	No connection
Spreadsheet calculation	3,34	0,07	2,14	2,20	2,99%	0,17	No connection
Database management	3,27	0,01	2,92	2,93	0,27%	0,05	No connection
SQL	0,43	0,12	1,51	1,63	7,23%	0,27	Weak
Programming	3,49	4,00	4,39	8,39	47,71%	0,69	Strong
OOP	0,07	0,00	0,37	0,37	0,81%	0,09	No connection

Table 17: How strong is the connection is between the grouping by genders in 11^{th} grade

Subject	$\bar{\bar{x}}$	S_k	S_b	S	H^2	H	Connection
Theoretical knowledge	14,34	3,09	7,53	10,63	29,12%	0,54	Middling strong
Word processing	6,29	0,02	2,80	2,81	0,67%	0,08	No connection
Spreadsheet calculation	3,16	0,05	2,14	2,18	2,17%	0,15	No connection
Database management	2,54	0,04	3,07	3,11	1,17%	0,11	No connection
SQL	0,35	0,06	1,24	1,30	4,95%	0,22	Weak
Programming	1,42	0,64	3,57	4,22	15,22%	0,39	Middling weak
OOP	0,14	0,00	0,60	0,60	0,12%	0,04	No connection

Table 18: How strong is the connection is between the grouping by genders in 12^{th} grade

According to the results the theoretical knowledge shows middling strong connection with gender in 11^{th} grade, because the gender explains knowledge in ˜39%. This value in the case of programming is ˜48%, which means strong connection

(Table 17.). We can declare here that our hypothesis, namely that boys are better in programming to be true.

We can see in 12^{th} grade that the earlier discussed two subjects (theoretical knowledge and programming) show a difference by gender (Table 18.). The gender explains theoretical knowledge in ˜29%. This value means middling strong connection, while in programming "just" in 15%, which means middling weak connection.

3. Summary

By calculated means of different subjects we can see specifications in computer science given by the National Curriculum are not totally followed. Word processing has been learned by most of the students, but spreadsheets calculation was generally neglected: ˜10% of the students have never learned that. The others could give correct answers on ˜30% of the questions. The situation is worse in other subjects, and object oriented programming was not taught at all in the secondary grammar school.

In the case of most subjects there is no observable difference in knowledge between the two genders.

The hypothesis in the introduction said boys get better marks in programming than girls. We know our hypothesis to be true now, and showed that the difference is growing in 11^{th} grade.

4. Proposal

The hypothesis was true: there is a big difference by gender in programming knowledge. This means at the same time that we should teach programming in another way for girls than for boys. Girls need other didactical methodes. There are already some working examples: there is an international Computer Science Bsc for girls at the Hochschule Bremen in Germany which is very successfull [6]. All of the students who qualify have a workplace after the school in the area of Computer Science. It would be not effectless to try a similar teaching method in Hungary too.

References

[1] Gábor Kiss - The Concept to Measure and Compare Students Knowledge Level in Computer Science in Germany and in Hungary / Acta Polytechnica Hungarica, 2008 Volume 5., pp:145.158, 2008, ISSN: 1785–8860.

[2] `http://nero.banki.hu`

[3] Varga Lajos: Kutatás-módszertan 151–156.

[4] Korpás Attiláné dr.: Általános statisztika II. 95–99.

[5] Korpás Attiláné dr.: placeGivenNameÁltalános middlenamestatisztika SnI. 152–153.

[6] http://www.hs-bremen.de/internet/de/studium/stg/ifi/