Research Article

Investigating factors affecting students' performance to PISA Science items

A. Psalidas*[1], C. Apostolopoulos[2] and V. Hatzinikita[3]

[1]*Athens College, General Lyceum,Athens, Hellas*
[2] *Varvakeio School, General Laboratory Lyceum,Athen, Hellas*
[3]*Associate Professor, Hellenic Open University, Educational Research and Pedagogy Unit,*
School of Humanities, Hellenic Open University, Hellas

_____

Abstract

The present paper aims to investigate, on the one hand, the extent to which PISA Science items validly assess the knowledge and skills of 15 year-old Greek students, while, on the other hand, to examine the effect of the following factors: student's gender, scientific processes and contexts (situations) on the students' performance in these PISA items. The research used paper-and-pencil test with published PISA Science items, conducted individual semi-structured interviews with 15 year-old students and finally marked the students' responses, according to the PISA marking guide. The basic finding resulting from the data analysis is that the paper-and-pencil test with the PISA Science items does not tend, unlike the interview, to effectively record the Greek students' Science knowledge and skills. Moreover, the analysis revealed that the performance of students in the PISA Science items (paper-and-pencil test and interview) tend to be independent of the student's gender and depend on the context in which the knowledge and processes are assessed. Additionally, the possible correlation between the students' performance and the factor of scientific processes seems to depend on the setting in which the students provide their responses (paper-and-pencil test or interview).

Keywords: PISA, scientific literacy, paper and pencil test, interview, students' performance.
_____

1. The Programme for International Student Assessment

The Programme for International Students Assessment (PISA) was launched in 1997 and is a major policy initiative in member countries of the Organisation for Economic Co-operation and Development (OECD) and in several non-member partner countries, which aims to regularly monitor the outcomes and the progress of their education systems in terms of 15 year-old students' achievement. The primary reason for developing and conducting this large-scale international assessment is to provide empirically grounded information which will inform policy decisions [1, 2, 3]. PISA aims to measure how well students, at the age of 15, are prepared to meet the challenges they may encounter in future life. PISA assesses domain – specific cognitive areas (Reading, Mathematical and Scientific literacy) not so much in terms of mastering the school curriculum but in terms of the important knowledge and skills needed in adult life. Emphasis is on the mastery of processes, the understanding of concepts and the ability to function in various situations within each assessment domain [3, 4]. PISA assessment takes place every three years and each cycle looks in depth at a major assessment domain, although it includes items from all three domains. In particular, the first cycle of PISA was conducted in 2000 (with a primary focus on Reading), the second in 2003 (with a primary focus on Mathematics) and the third in 2006 (with a primary focus on Science), while the programme is expected to be completed in 2015.

Age 15 is chosen because at this age, in most OECD countries, students are normally near the end of their compulsory schooling, and so, some useful indication of the knowledge, skills and attitudes acquired over approximately ten years of this initial period of basic schooling is gained from an assessment at that stage.

2. The PISA Definition of Scientific Literacy

The key-concept of the PISA theoretical framework as well as an objective of the compulsory education, according to PISA, is the concept of literacy, which is concentrated in three domains, namely Reading, Mathematical and Scientific Literacy.

According to PISA[1], scientific literacy is "the capacity to use scientific knowledge, to identify questions and to draw

_____
[1] Because the present research was conducted in 2005-06 it provides the definition of the scientific literacy of the 2000 and 2003 PISA cycles.

evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity." [1 (p. 60), 4 (p. 133)]. It should also be noted that PISA does not categorise students as either scientifically literate or scientifically illiterate. Instead, there is a progression from the less developed to the more developed scientifically literate students. According to the Education Research Centre of Greece[2] [5 (p. 330)], scientifically literate are considered the students who:

- know and understand the scientific concepts and processes necessary for their present and future participation in society,
- are able to ask questions and provide responses to issues emerging from their daily experiences and attracting their interest,
- are able to describe, explain and predict natural phenomena,
- are able to read and understand popularised scientific articles in the press and discuss their validity,
- are able to identify the scientific aspect of an issue, when they are asked to decide (at national and local level), and express scientifically and technologically documented aspects,
- are able to assess scientific information with respect to both their source and the methods followed for its generation,
- are able to develop and assess arguments based on specific data, as well as implement the appropriate conclusions in practice.

The PISA Programme defines three[3] interrelated organizational dimensions structuring and operationally defining literacy, in general, as well as its separate domains (Reading, Mathematical and Scientific Literacy). In the case of scientific literacy, on which the present paper is focused, these three dimensions are as follows [1, 4]:

- Scientific knowledge or concepts (which will be assessed by applying it to specific subject matter);
- Scientific processes (which, because they are scientific, will involve knowledge of Science, although in the assessment this knowledge should not form the major barrier to success);
- Context or Situations (in which the knowledge and processes are assessed and which take the form of Science-based issues).

The above mentioned dimensions are operationally defined as follows [4].

- Scientific concepts. The major scientific themes for assessing scientific literacy are: structure and properties of matter, atmospheric changes, chemical and physical changes, energy transformations, forces and movements, form and function of organisms, human biology, physiological changes, biodiversity, genetic control, ecosystems, the Earth and its place in the universe and geological changes.
- Scientific processes. Processes are mental (and sometimes physical) actions used in conceiving, obtaining, interpreting and using evidence or data to gain knowledge or understanding. The PISA 2003 scientific processes were:
  - Describing, explaining and predicting scientific phenomena. In this process students demonstrate their understanding by applying appropriate scientific knowledge in a given situation. It involves describing or explaining phenomena and predicting changes and may also involve recognising or identifying appropriate descriptions, explanations and predictions.
  - Understanding scientific investigation. It involves recognising scientifically investigate questions or suggesting a question that could be scientifically investigated in a given situation. It also involves identifying or recognising evidence needed in a scientific investigation: for example, what things should be compared, what variables should be changed or controlled, what additional information is needed or what action should be taken so that relevant data can be collected.
  - Interpreting scientific evidence and conclusions. This means making sense of scientific findings as evidence for claims or conclusions. It may involve accessing scientific information and producing and communicating conclusions based on scientific evidence. It may also involve selecting from and communicating alternative conclusions in relation to the evidence; giving reasons for or against a given conclusion in terms of the data provided or identifying the assumptions made in reaching a conclusion, and reflecting on and communicating the societal implications of scientific conclusions.
- Context: It requires application of the selected scientific knowledge and the use of scientific processes in important situations reflecting the real world and involving ideas of Science. The areas of application in which the scientific knowledge and processes are assessed are:
  - Life and Health (health, disease, nutrition, maintenance and protection of biodiversity, natural and biological systems and their interdependence).
  - Earth and Environment (pollution, production and loss of soil, weather and climate).
  - Technology (biotechnology, use of materials and waste disposal, energy consumption and management, transportation).

3. The format of the PISA items

The PISA items are structured in test-units. Each test-unit comprises an introductory stimulus-material (a written

passage, or a text accompanying a table, chart, graph, or a diagram, extracts of scientific and popularised newspaper articles, historical documents, research protocols or a combination of the above), plus the items which are a set of independently scored questions of various types (multiple-choice, short answers, and extended responses), which require finding and exploiting the information presented in the stimulus-material in order to be answered. Each unit refers to a real situation of everyday life.

4. Relevance and Aims of the Research

The international comparative student assessment surveys adopt the paper-and-pencil test as the tool for collecting data. However, the relevant literature raises questions regarding the validity of the paper-and-pencil test in recording the students' actual knowledge and skills particularly in Science [6-10].

The investigation of the validity of a paper-and-pencil test consists in verifying the extent to which the students' performance in this test actually assesses their knowledge. The most appropriate tool for investigating this issue is the interview, as it allows further explanations and clarifications of the students' responses and, as a result, demonstrates and records their knowledge and way of thinking [6, 8, 9]. The research works that adopted a similar reasoning revealed that the students' overall performance, in the framework of verbal expression (interview), is much better than the respective overall performance of the same students in written expression (paper-and-pencil tests) [6, 7, 9, 10].

The overview of the relevant literature showed that similar methodological issues have been investigated with respect to other international student assessments in Science, e.g. TIMSS [6, 9], but have not been investigated with respect to PISA, at least regarding research papers published in the international scientific journals in English [11].

In the direction of this reasoning, the present paper aims to partially fill the void on this area, as its main aim is to investigate the extent to which the written responses of 15 year-old Greek students to PISA Science items actually record their Science knowledge and skills. Furthermore, the paper aims at investigating the extent to which the students' performance in the paper-and-pencil test and the interview, as regards the same PISA Science items, depends on: a) gender (male, female), b) scientific process (Describing, explaining and predicting scientific phenomena, Understanding scientific investigation, Interpreting scientific evidence and conclusions) and c) context (Life and Health, Earth and Environment, Technology).

In addition, the present paper intends to contribute to an essential discussion about the results of the Greek students in PISA, as their performance[4] was poor in all three PISA cycles, in comparison with the other participating countries, while the discussion about this issue in Greece is almost exclusively monopolized by the mass media, tending to be focused only on commenting and being completely engrossed in the low rank of the country. The latter is connected with the particularly limited systematic research carried out on the results of the Greek students in PISA [11,

12, 13, 14].

5. Research Method

Research Process
A research process comprising the following stages was prepared for the needs of the research:

- The students complete a paper-and-pencil test with PISA items.
- The students' responses to the paper-and-pencil test are marked according to the relevant PISA marking guides. The students' scores depict their overall performance and are used for investigating the dependence of their performance on the factors: gender, scientific process and context.
- Individual semi-structured interviews are carried out, with selected students of the sample in the same items they confronted in the paper-and-pencil test.
- The students' responses to the interviews are marked according to the relevant PISA marking guides. The marked students' responses to the interviews intend both to record their performance in the same items, though in a verbal communication setting, and to investigate the dependence of their performance in the interview on the gender, the scientific process and the context.
- The students' performance in the paper-and-pencil test is compared with their performance in the interview so as to realise (i) the extent to which the students' responses to the paper-and-pencil test record the knowledge they actually have and (ii) the dependence of these differences on the factors: gender, scientific process and context.

The Sample
The research sample consisted of 94 students at the age of 15 (48 boys and 46 girls), first-grade students in a lyceum during academic year 2005-2006 (10th grade).

Data Collection Tools
(A) The paper-and-pencil test
The paper-and-pencil test used in the present research included, among the available[5] PISA test-units, six test-units those which are more proximate to Physics. Among these items, two of them ("Semmelweis' Diary" and "Daylight") were used in previous PISA cycles (PISA 2000 and PISA 2003, respectively), while the other four were proposed by PISA as indicative of its approach. More specifically, the test included a total of 14 items.

The titles of the test-units, the number of items included, the scientific process that examines each item and the context in which each item is incorporated are presented in the following Table 1.

The maximum duration for completing the test of the research was 30 minutes, which was determined by analogy with the maximum time needed for completing the original PISA test.

---

[4] The average performance of Greek students in scientific literacy was 461 score-points in PISA 2000 (25th among 30 countries, OECD) [15], 481 score-points in PISA 2003 (30th among 40 countries) [16] and 473 score-points in PISA 2006 (38th among 57 countries) [17].

---

[5] The PISA test-units are classified. Only a very limited number of test-units are published after the end of each PISA cycle.

Table 1:

Test – units, number of items, scientific processes and contexts

Test-unit	Number of items	Scientific process	Context
«Climate Change»	1st	Interpreting scientific evidence and conclusions	Earth and Environment
«Daylight»	1st	Describing, explaining and predicting scientific phenomena	Earth and Environment
	2nd	Describing, explaining and predicting scientific phenomena	Earth and Environment
«Semmelweis' Diary»	1st	Interpreting scientific evidence and conclusions	Life and Health
	2nd	Interpreting scientific evidence and conclusions	Life and Health
	3rd	Describing, explaining and predicting scientific phenomena	Life and Health
	4th	Describing, explaining and predicting scientific phenomena	Life and Health
«Corn»	1st	Describing, explaining and predicting scientific phenomena	Life and Health
	2nd	Describing, explaining and predicting scientific phenomena	Life and Health
	3rd	Understanding scientific investigation	Earth and Environment
«Calf Clones»	1st	Interpreting scientific evidence and conclusions	Life and Health
	2nd	Describing, explaining and predicting scientific phenomena	Life and Health
«Buses»	1st	Describing, explaining and predicting scientific phenomena	Technology
	2nd	Interpreting scientific evidence and conclusions	Earth and Environment

As regards completing and coding the paper-and-pencil test, the processes are as follows:

Before the paper-and-pencil test was handed out and completed by the sample students:

- the students were given instructions about its completion and the teachers were given instructions about the process, in accordance with the PISA guides,
- 94 booklets were prepared, including the 6 test-units in 23 different arrangements, so that the students' responses could not be affected by the order of the items, and
- the sample students were informed about the purpose and the aim of the research both orally (by

the researchers) and through printed material distributed to them.

The test was completed by the students and their responses were subsequently codified and marked, according to the PISA marking guides. Two independent researchers codified and marked the students' responses. In case the marks of the two researchers varied, the final mark was the average of the two marks. Actually, there were very few differences between the marks of the two researchers.

(B) The interviews

The data collected from the test booklets were validated through in-depth interviews with a representative sample of the respondents. In particular, 20 students (10 boys and 10 girls), corresponding to 21.3% of the sample that completed the paper-and-pencil test [9], participated voluntarily in the individual semi-structured interviews. The 20 students participating in the interviews were selected through stratified random sampling. The students' marks in the test were classified in ascending order, were distributed in 4 classes of similar frequencies, the number of interviews that had to be received from each class was determined, and then the male and female students who were going to be interviewed were randomly selected by the volunteers. The selection was made using the Microsoft Excel generator of random numbers.

There was an interval of five weeks between the paper-and-pencil tests and the interviews (two of these weeks were the Christmas holidays) so that:

- the items of the test were not very recent to the students,
- the students did not acquire any further school knowledge on the items of the test and
- the appropriate time was allotted to the two raters to complete marking.

Individual semi-structured interviews were conducted because this kind of interview allows focusing on the students' way of thinking and on the in-depth investigation of their knowledge and skills. The process of the interview was as follows: the interviewer read out each item of the test and then invited the student to read out their written response. Next, after informing the student that they could (if necessary) change their response, the interviewer asked for documented explanations about the formation or selection of their response. In case of an open-constructed item, the student was invited to clarify and explain their response. In case of a closed-constructed item, the student was invited to document their response and justify why the other options were rejected.

The interviews were conducted at school and each interview lasted as long as one teaching hour (40 minutes).

All the interviews were tape-recorded with the students' consent, were transcribed and then the responses were marked according to the PISA marking guides.

Analysis Method

The variables that were compared came either from independent groups (performance of boys and performance of girls) or from the same group (performance of the same student in every scientific process and context). In all cases the Kolmogorov-Smirnov Normality Test was performed in advance [18]. The comparison between the average values of

variables coming from the independent group was made using the t-test. In case the variables came from the same group, the comparison was made using the paired t-test when two variables were compared and the Friedman test, when three variables were compared, as they did not all follow the normal distribution. Whenever the Friedman test expressed a statistical significance, the variables were compared in pairs using the Wilcoxon signed-rank test (on acceptance of the Bonferroni criterion) aiming at detecting the pair or pairs whose performance differences were statistically significant. The level of statistical significance selected for all the comparisons was the usual 0.05 (5%). All the tests were conducted using the Statistical Package for Social Sciences (SPSS) software, Version 13.0.

6. Results

The results of the analysis of the students' responses in the paper-and-pencil test (N=94, see Table 2) are as follows:

- The students' average percentage performance (A%P) in the paper-and-pencil test is 46.12 (or 9.22 in the 20-grade scale).
- The average percentage performance of boys' (49.57) is better than that of the girls (42.52). However, this difference in the performance between boys and girls is not statistically significant (t-test, p=0.06).
- The students' average percentage performance in contexts of "Technology", "Life and Health" and "Earth and Environment" is 60.6, 54.5 and 41.1 respectively. These differences in average percentage performance are statistically significant (Friedman, p=0.002). After comparing the students' performance in the different pairs of contexts it was realised that the students deliver a statistically remarkably better performance (a) in the context of "Technology" in comparison with the context of "Earth and Environment" (Wilcoxon, p=0.000) and (b) in the context of "Life and Health" in comparison with the context of "Earth and Environment" (Wilcoxon, p=0.000). On the other hand, the difference in the average percentages between the contexts of "Technology" and "Life and Health" is not statistically significant (Wilcoxon, p=0.359).
- The students' average percentage performance in the scientific process of "Understanding scientific investigation" (48.9) is better than the respective percentage of their performance in the scientific process of "Describing, explaining and predicting scientific phenomena" (46.5), which in turn is better than the respective performance in the scientific process of "Interpreting scientific evidence and conclusions" (45.3). However, these differences are not statistically significant (Friedman, p=0.849).

The results in the paper-and-pencil test of the 20 students participating in the interview appear in Table 3. These students achieved a slightly better average performance than that achieved by the overall sample.
According to the results obtained from the analysis of the students' responses in the interview (N=20, see Table 4):

- The students' average percentage performance in the interview is 76.18 (or 15.24 in the 20-grade scale).

- The average percentage performance of boys' (78.24) is better than that of the girls (74.12), though the difference in their performance is not statistically significant (t-test, p=0.602).
- The students' average percentage performance in the context of "Technology" (85.0) is better than the respective percentage performance in the context of "Earth and Environment" (81.4), which in turn is better than the respective percentage performance in the context of "Life and Health" (71.1). These differences in the average percentage performances are statistically significant (Friedman, p=0.008), whereas the differences between the pairs are not statistically significant (Wilcoxon, $p_{1,2}=0.05$, $p_{1,3}=0.375$ and $p_{2,3}=0.141$).
- The students' average percentage performance in the scientific process of "Understanding scientific investigation" (85.0) is better than the respective percentage performance in the scientific process of "Interpreting scientific evidence and conclusions" (82.1), which in turn is better than the respective percentage performance in the scientific process of "Describing, explaining and predicting scientific phenomena" (70.6). The differences among the three above mentioned variables are statistically significant (Friedman, p=0.007), whereas the differences detected between the pairs are not statistically significant (Wilcoxon, $p_{1,2}=0.553$, $p_{1,3}=0.023$ [>0,050/3] and $p_{2,3}=0.176$).

Table 5 presents briefly the differences (improvements) in students' performance between the paper-and-pencil test and the interview with respect to the investigated variables. The comparative analysis of these differences shows that:

- The students' average percentage performance from the paper-and-pencil test to the interview improved by 26.18, a statistically significant improvement (paired t-test, p=0.000). In other words, the initial paper-and-pencil performance improved through the verbal form of communication, by 52.4%.
- The average improvement in the percentage performance of boys from the paper-and-pencil test to the interview (27.06) is greater than the respective improvement of girls (25.29), although this difference is not statistically significant (t-test, p=0.796).
- The students' average percentage performance from the paper-and-pencil test to the interview showed a greater improvement in the context of "Earth and Environment" (35.7), a smaller improvement in the context of "Technology" (15.0) and an even smaller improvement in the context of "Life and Health" (14.2). Besides, these differences are statistically significant (Friedman, p=0.007). Among the three pairs compared the only statistically significant difference is noted in the pair "Earth and Environment" and "Life and Health" (Wilcoxon, p=0.005).
- The students' average percentage performance from the paper-and-pencil test to the interview showed a greater improvement in the scientific process of "Interpreting scientific evidence and conclusions" (31.4), a smaller improvement in the scientific process of "Understanding scientific investigation" (25.0) and an even smaller improvement in the scientific process of

"Describing, explaining and predicting scientific phenomena" (23.3). However, the above mentioned differences among the three variables are not statistically significant (Friedman, p=0.111).

7. Conclusions

The main finding of the present research paper is that the PISA Science items that were used do not seem to efficiently record the students' Science knowledge and skills because of the very clear differences between the outcomes in the two settings (paper-and-pencil test and interview). Nevertheless, this finding, which is the same as in other research works that investigated Science items of the TIMSS international study in exactly the same way [6, 9], should not be interpreted as an element improving the rank of our country in the scientific literacy of the PISA 2000 and 2003 cycles.

Furthermore, apart from the finding that the performance achieved by the sample-students in the paper-and-pencil test of the present research is as poor as that of the Greek students in the PISA 2000 and PISA 2003 cycles,

the present research has highlighted the effects produced by specific factors on the students' performance:

(a) The students' performance in the paper-and-pencil test and the interview as well as the improvement in their performance from the paper-and-pencil test to the interview does not depend on the gender of the students asked.

(b) The students' performance in both the paper-and-pencil test and the interview as well as the improvement in their performance from the paper-and-pencil test to the interview depends on the context of the items.

(c) The students' performance tends to depend on the scientific process examined by the items exclusively in the framework of the students' participation in the interview.

Nevertheless, the above mentioned findings cannot be generalised in Greek students, as the research sample was not representative of 15 year-old Greek students. Therefore, the repetition of the research in a representative sample of 15 year-old Greek students would be particularly useful as long as the used tool is at the same time enriched with Science test-units published in the framework of the PISA 2006 cycle.

Table 2:

Students' performance in the paper-and-pencil test (N=94)

	Total performance	Gender		Context			Scientific process		
		Boys	Girls	Earth and Environment	Life and Health	Technology	Interpreting scientific evidence and conclusions	Understanding scientific investigation	Describing, explaining and predicting scientific phenomena
Average percentage performance	46.12	49.57	42.52	41.11	54.45	60.64	45.29	48.94	46.45
Significance of differences on average percentage performance		Not statistically significant (t-test, p=0.06)		Statistically significant (Friedman, p=0.002) — Statistically significant (Wilcoxon, p=0.000) — Statistically significant (Wilcoxon, significant p=0.000) — Not statistically significant (Wilcoxon, p=0.359)			Not statistically significant (Friedman, p=0.849)		

Table 3:

Performance in the paper-and-pencil test of the students who participated in the interviews (N=20)

	Total performance	Gender		Context			Scientific process		
		Boys	Girls	Earth and Environment	Life and Health	Technology	Interpreting scientific evidence and conclusions	Understanding scientific investigation	Describing, explaining and predicting scientific phenomena
Average percentage performance	50.00	51.18	48.82	45.73	56.88	70.00	50.72	60.00	47.22
Significance of differences on average percentage performance		Not statistically significant (t – test, p=0.814)		Not statistically significant (Friedman, p=0.086)			Not statistically significant (Friedman, p=0.534)		

Table 4:

Students' performance in the interview (N=20)

	Total performance	Gender		Context			Scientific process		
		Boys	Girls	Earth and Environment	Life and Health	Technology	Interpreting scientific evidence and conclusions	Understanding scientific investigation	Describing, explaining and predicting scientific phenomena
Average percentage performance	76.18	78.24	74.12	81.43	71.11	85.00	82.14	85.00	70.56
Significance of difference in average percentage performance		Not statistically significant (t – test, p=0.602)		Statistically significant (Friedman, p=0,008) — Not statistically significant (Wilcoxon, p=0.050); Not statistically (Wilcoxon — significant p=0.375); Not statistically significant (Wilcoxon, p=0.141)			Statistically significant (Friedman, p=0.007) — Not statistically significant (Wilcoxon, p=0.553); Not statistically (Wilcoxon, — significant p=0.023>0,050/3); Not statistically significant (Wilcoxon, p=0.176)		

Table 5:

Performance differences between the paper-and-pencil test and the interview (N=20)

	Difference in total average percentage performance	Gender		Context			Scientific process		
		Boys	Girls	Earth and Environment	Life and Health	Technology	Interpreting scientific evidence and conclusions	Understanding scientific investigation	Describing, explaining and predicting scientific phenomena
Difference in average percentage performance	26.18	27.06	25.29	35.71	14.24	15.00	31.42	25.00	23.33
Significance of difference in average percentage performance	Statistically significant (paired t-test, p=0.000)	Not statistically significant (t-test, p=0.796)		Statistically significant (Friedman, p=0.007) — Statistically significant (Wilcoxon, p=0.005); Not statistically (Wilcoxon, — significant p=0.085); Not statistically significant (Wilcoxon, p=0.341)			Not statistically significant (Friedman, p=0.111)		

References

1. OECD, Measuring Student Knowledge and Skills. A New Framework for Assessment, OECD, Paris (1999).
2. OECD, Reading for Change. Performance and Engagement across Countries. Results from PISA 2000, OECD, Paris (2002).
3. OECD, Assessing Scientific Reading and Mathematical Literacy. A Framework for PISA 2006, OECD, Paris, (2006).
4. OECD, The PISA Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills, OECD, Paris, (2003).
5. Education Research Centre, PISA: Programme for International Student Assessment, Eptalofos, Athens, (2007).
6. A.Harlow, and A.Jones, "Why students answer TIMSS science test items the way they do. Research in Science Education", 34, 221, (2004)
7. P.Murphy, "The IEA Assessment of Science Achievement. Assessment in Education" 3(2), 213, (1996).
8. S.Messick, "Meaning and Values in Test Validation: The Science and Ethics of Assessment", Educational Researcher, 18, 5, (1989).
9. J.Schoultz, R.Säljö, and J.Wyndhamn, "Conceptual knowledge in talk and text: What does it take to understand a science question?", Instructional Science, 29, 213, (2001).
10. P.Warwick, R.S.Linfield, and P.Stephenson, "A Comparison of Primary School Pupils' Ability to Express Procedural Understanding in Science through Speech and Writing", International Journal of Science Education, 21(8), 823, (1999).
11. V.Hatzinikita, K.Dimopoulos, and V.Christidou, "PISA test items and school textbooks related to science: A textual comparison", In Press in Science Education.
12. K.Apostolopoulos, "Investigating factors affecting students' responses to PISA Science items", Unpublished Postgraduate Dissertation, Hellenic Open University, Patras, (2007).

13. Dimopoulos, K., Hatzinikita, V., & Christidou, V. (2005). Textes et messages pedagogiques en sciences physiques: une approche interpetative possible des resultats de enquete PISA. In: A. Giordan, J. L. Martinand, & D. Raichvarg (eds), *Actes des XIXes Journales Internationales sur la Communication, l' Education et a Culture Scientifiques, Techniques et Industrielles. Par les mots et par les choses*. France: Universite Paris VI.

14. A.Psalidas, "Why male/female students respond to PISA Science items the way they do", Unpublished Postgraduate Dissertation, Hellenic Open University, Patras, (2007).

15. OECD, Knowledge and Skills for Life. – First Results from PISA 2000. OECD, Paris, (2001).

16. OECD, Learning for Tomorrow's World – First Results from PISA 2003. OECD, Paris, (2004).

17. OECD, PISA 2006: Science Competencies for Tomorrow's World, Volume 1 – Analysis, OECD, Paris, (2007).

18. A.Katsis, G.Sideridis and A.Emvalotis, "Statistical methods in social sciences"(in press).