

Conceptual understanding versus algorithmic problem solving: Further evidence from a national chemistry examination

Dimitrios Stamovlasis, Georgios Tsaparlis*, Charalambos Kamilatos,
Dimitrios Papaioikonomou, and Erifyli Zarotiadou

University of Ioannina, Department of Chemistry, GR-45 10 Ioannina, Greece
e-mail: gtseper@cc.uoi.gr

Received 12 April 2002, accepted 2 March 2005

Abstract: Following our previous paper (*Chem. Educator*, **2004**, *9*, 398-405), we analyze further the results of a national examination from the perspective of conceptual learning versus algorithmic problem solving. Detailed achievement data were studied for a sample of 499 eleventh-grade students (age about 17), who were following various branches or streams leading to all kinds of higher-education studies in Greece (the 'Positive', the 'Theoretical', and the 'Technological' Branches). Using qualitative criteria, we distinguished the questions into: (i) simple knowledge-recall, (ii) conceptual, and (iii) well-practiced (algorithmic), stoichiometric, exercises. The latter could further be divided into simple and more demanding ones. As in the previous study, this categorization was also supported by statistical principal component analysis, but this time a marginal structure was extracted, because (possibly) of the limited number and the low difficulty of the postulated conceptual questions. The interest of the study lies mainly in the comparison among the different branches, with the students of the Positive Branch demonstrating the highest mean scores. In addition, students' thinking was categorized according to Nakhleh's scheme. The Positive Branch had the highest number of students with algorithmic and with conceptual ability, but all branches had about equal share of students high only in conceptual ability. [*Chem. Educ. Res. Pract.*, 2005, **6** (2), 104-118]

Keywords: *conceptual understanding; algorithmic problem solving; national examinations; principal components analysis; Nakhleh's scheme.*

Introduction

School chemistry, as any other school subject, is an internationally more or less settled subject with respect to its content and objectives. In upper secondary school, chemistry is usually presented as a simplified and reduced version of college general chemistry. Researchers in science education contend that such a situation and process is not correct: in school we do not aim at educating future scientists, but to develop instead a scientific literacy which is sufficient to make the future citizen able to understand and participate in decision making on crucial social and economic matters (*National Education Standards*, 1993). A main feature of the way school chemistry is taught and tested all over the world is that the emphasis is often placed on learning rules and algorithms that enable students to respond with success to examination questions, including relatively complicated computational 'problems'/exercises. What happens, however, in the case of conceptual questions, even apparently simple ones? Further, are all students, irrespective of their aspirations for higher studies, demonstrating the same interest, inclination, and ability in the various kinds of test questions?

In a recent previous paper (Stamovlasis *et al.*, 2004), we analyzed the results of the Greek

Chemistry Education Research and Practice, 2005, **6** (2), 104-118

National Examination from the perspective of conceptual understanding versus algorithmic problem solving. Detailed achievement data in the special subject 'Chemistry for the Positive Branch' (see below) were studied for a sample of 647 eleventh-grade students (age about 17) who were oriented toward science, engineering or medical subjects (following the *Positive Branch* or *Stream*). It was demonstrated that principal component analysis (PCA) could serve as a tool for scrutinizing the items of examination papers in chemistry. Further, national, large-scale examinations provide reliable data that are appropriate for such an analysis. PCA distinguished between conceptual questions and the computational, well-practiced (algorithmic) questions. Some more demanding computational questions (requiring analysis and synthesis) shared some common space with the conceptual questions. On the other hand, the easy recall and simple application-of-knowledge questions were separated out from all other questions. The above conclusions were also supported by Multivariate Analysis of Variance (MANOVA). Achievement was at about the same level in the conceptual and the more demanding algorithmic questions. Finally, a scheme suggested by Nakhleh (1993) was also used to categorize the students according to the various categories of algorithmic versus conceptual thinking.

In this paper, we carry further the analysis of the results of the Greek National Examination in Greece from the same perspective, this time by considering the course 'Chemistry for General Education', taken by all eleventh-grade students (age about 17), irrespective of their inclination for specific subjects. The content of the course was mainly from organic chemistry. We repeat here that this examination was the first given after an educational reform in which, for the first time in the past thirty years or so, some of the questions required some form of conceptual understanding; in former cases, the dominant character of the examination questions was equally distributed between knowledge recall and algorithmic exercises. Both of these abilities (recall and algorithmic) were well practiced both within and outside of school. In contrast, the students did not have previous special training in manipulating conceptual questions in the specific domain of organic chemistry.

Formal statistical data revealed that 'Chemistry for General Education', among nineteen general and specialized nationally tested subjects (the latter including 'Chemistry for the Positive Branch') was the subject with the largest difference between oral and written marks, the former being allocated by the class teachers, while the latter coming from the objective national examination: a 30.3% difference. This was apparently due to the difficulty of the chemistry paper.

The rationale for dividing examination questions into conceptual and algorithmic/computational ones, together with the categorization scheme of Nakhleh (1993), which is also used in this paper, have been discussed and reviewed in our previous paper (Stamovlasis et al., 2004). Here we only need to repeat and stress three points: (a) that various authors have used various methods to categorize questions as being algorithmic or as requiring conceptual understanding; (b) conceptual questions have been associated with higher-order cognitive skills (HOCS), and algorithmic questions with lower-order cognitive skills (LOCS) (Zoller, et al., 1995; Zoller & Tsaparlis, 1997); (c) the degree to which an examination item is categorized as requiring conceptual thinking is, to some extent, a function of the students' background and the sort of teaching they have been exposed to in class (Niaz, 1995).

A question that requires just LOCS for some students may require a shift to HOCS for others in a different context. It is therefore possible that questions (or some of them) that are categorized here as conceptual might be considered by students with a different background from that of the ones in our study not as conceptual, but as requiring just knowledge. On the other hand, computational questions may require for their answer not just the use of algorithms, but also conceptual understanding and critical thinking. As such, their relation

with conceptual questions may be not dichotomous (Niaz, 1995). In this work, of necessity, questions were identified as conceptual according to the operational definition of Zoller and Tsaparlis (Zoller & Tsaparlis, 1997; Tsaparlis & Zoller, 2003). This assignment was further checked by proper statistical analysis (see below).

Method

The sample

Our sample consisted of 499 students (age about 17), who took the Greek National Examinations for the Eleventh Grade. All the students came from various urban schools and were representative of the student population in urban areas in Greece. The large size of the sample makes it very likely that it was a homogeneous sample, even if students came from many different schools. The examinations took place at the end of the school year in June 1998-99, and were part of the university placement examinations that started in eleventh grade and were completed at the end of twelfth grade. Chemistry (organic chemistry) was one of the examined subjects for all students.

All students took the same courses up to the end of tenth grade. Starting at eleventh grade, students had to follow one of three *branches* (or *streams*): The 'Positive' Branch (PB), the 'Theoretical' Branch (ThB), and the 'Technological' Branch (TB). The PB is for students who want to study science, engineering and related applied subjects, or medicine, or related subjects. The ThB is for students who want to study literature, law, humanities etc. Finally, the TB leads to the same studies as the PB, except for medical subjects, but attracted weaker students.

At the outset, it must be noted that the weakest students joined the TB, while the academically stronger students separated into the two other branches: those with an interest in the humanities, etc. went into the ThB and those with an interest in science, medicine and engineering into the PB. These latter ones got extra instruction in chemistry, above that received by the other two branches (see Stamovlasis *et al.*, 2004). Note that many schools had only two branches (ThB and PB) running together, while in a smaller number of schools all three branches were running.

All students took the same end-of-year examination in (organic) chemistry, and it is the results of this examination that are analyzed in this paper. In addition, the students in the PB took a further examination in (general/physical) chemistry, assessing the further instruction they received (the results of this examination have been analyzed in Stamovlasis *et al.*, 2004).

The numbers of students in our sample were distributed as follows: PB, $N = 234$, ThB, $N=172$, and TB, $N = 93$.

The test

The organic chemistry paper consisted of four parts: Part 1 contained largely recall, fixed-response questions, while Part 2 included both knowledge and a couple of simple conceptual questions (questions 2.3.a and 2.3.b). Parts 3 and 4 were more difficult, both involving stoichiometric calculations. The test and the marks allocated to each question and sub-question are included in Appendix 1. The test was constructed by a special committee of the Greek Ministry of Education. The present authors were not involved in the examination construction procedure, but two were involved in the grading procedure. Note that the national examination of Greece is similar to the examinations used for student selection for higher education, such as SAT (in the U.S.) and GCE (in England and Wales).

To facilitate analysis, a further distinction of the questions was made after agreement among the researchers. In Part 1, we grouped together the multiple-choice recall questions

1.1-1.4, and the fill-in questions on organic reactions 1.6 and 1.7, but we kept separate the open (knowledge-recall) question 1.5, and the one-to-one correspondence (monomers-polymers) 1.8. In Part 2, question 2.1 was of the right-wrong type, with explanation, while question 2.2 asked for two organic reactions; of particular interest is question 2.3, consisting of two conceptual sub-questions 2.3.a and 2.3.b which deal with the exhaust gases of cars with or without catalytic converters. In Part 3, questions 3.a and 3.b demanded numerical stoichiometric calculations on the fermentation of sugar (in grape must) to ethanol (chemical equation given). Question 3.c provides the second example of a simple conceptual question: students had to judge and explain whether the mass of produced wine would be different from that of the must. Finally, in Part 4 all three questions (4.a, 4.b, and 4.c) involved demanding stoichiometric computations that also required knowledge of organic reactions.

An alternative, independent/objective evaluation/categorization of the questions of the test can be done by using statistical *principal component analysis* (PCA). This can lead to a reduction in the number of variables and to the detection of structure in the relationships between variables (Anderson, 1984; Stamovlasis *et al.*, 2004). Using Kaiser's criterion (Kaiser, 1958) and the scree test (Catell, 1966) we arrived at a marginal structure, which is in agreement with our qualitative categorization. The marginal structure may be attributed to the limited number and the low difficulty of the postulated conceptual questions. The results of this analysis are given in Appendix 2.

The duration of the examination was three hours. The papers that supplied the data for this study were marked by two chemistry teachers who are among the authors of this paper (E.Z. and D.P.). The marking procedure was similar to the one reported in our previous study (Stamovlasis *et al.*, 2004).

Research questions

The following research questions were asked:

1. Was the postulated categorization of questions (conceptual versus algorithmic) supported by the data?
2. How were students, *assigned* as conceptual thinkers, distributed amongst the various branches? Were we right to assume that students with both high conceptual and algorithmic thinking would be found mainly in the PB? Were we right to assume that students with algorithmic thinking were expected to be distributed mainly in the PB and the TB?
3. Is Nakhleh's scheme operating in our case?
4. Is competence in algorithmic problem solving connected with competence in conceptual problem solving (and vice versa)?

In addition, we wanted to see what differences in achievement were to be found among the three branches. We expected, of course, that the PB would have the highest achievement.

Results and Discussion

Achievement

The mean score of the whole sample ($N = 499$) in the whole paper was about 50% [49.4%, with standard deviation (SD) 26.6%]. This should be contrasted with the mean score of the PB in the special advanced examination in general chemistry ($N = 647$, $M = 68.0\%$, $SD = 26.2\%$ - see Stamovlasis *et al.* 2004), demonstrating the distinctive features of the organic chemistry paper. A steady decline of mean scores is observed in going from Part 1 of the test (80.4%, SD 17.0) to Part 2 (50.0%, SD 29.6), to Part 3 (36.1%, SD 37.1), and to

Table 1. Mean scores (%) in the questions of the test that are of greatest interest to this study (standard deviations in parentheses).

Question	Total Sample (<i>N</i> = 499)	Positive Branch (<i>N</i> = 234)	Theoretical Branch (<i>N</i> = 172)	Technological Branch (<i>N</i> = 93)
2.3.a	40.4 (45.2)	52.9 (46.0)	31.4 (42.2)	25.9 (40.5)
2.3.b	54.5 (42.6)	60.0 (42.5)	51.4 (42.8)	46.4 (40.7)
3.a	44.3 (44.0)	60.6 (43.2)	33.0 (40.6)	24.2 (37.3)
3.b	31.2 (42.6)	47.1 (46.0)	20.9 (36.5)	10.2 (26.2)
3.c	29.7 (43.6)	37.9 (46.9)	22.8 (39.5)	22.0 (38.7)
4.a	31.9 (42.4)	42.3 (46.4)	26.9 (38.2)	15.1 (30.8)
4.b	34.4 (43.0)	46.0 (45.7)	28.9 (39.6)	15.5 (32.4)
4.c	21.5 (33.1)	31.9 (38.3)	15.4 (26.6)	6.5 (18.2)

Part 4 (29.0%, SD 35.2). As expected, the PB had a highest mean score (59.0%, SD 27.5), while intermediate was the ThB (44.1%, SD 23.7) and lowest the TB (35.1%, SD 19.4). The differences are statistically significant, as shown by one-way ANOVA for independent samples.

Table 1 contains detailed mean score results on the various questions that were of the greatest interest to this work. One-way ANOVA for dependent samples gives (through the Tukey test) the following critical values of differences of means for statistical significance at $p = 0.05$: 8.6 for the total sample, 13.3 for the PB, 13.4 for the ThB, and 16.3 for the TB.

The two simple conceptual questions 2.3.a and 2.3.b proved hard for many students. More demanding was 2.3.a (40.4%), because of the chemical equation, while in 2.3.b the mean score was moderate (54.5%).

The mean scores were low in questions 3.a and 3.b: 44.3% versus 31.2%. Question 3.b required 3.a for its answer, and that is the reason for the much lower scores in 3.b than in 3.a. Question 3.c provides an example of a question that could be treated either through stoichiometric calculations (by using the previously obtained masses of alcohol and sugar from questions 3.a and 3.b) or as a simple qualitative question. Recall that the question asked students to decide and explain whether the mass of produced wine would be different from that of the grape must. Working from the chemical equation, given that one of the products (carbon dioxide) is a gas, it is an easy qualitative conclusion that the mass of the produced wine is less than that of the must. A possible explanation for the low mean mark (29.7%) could be that many students had difficulty carrying out the stoichiometric calculations or to draw the qualitative inference and reach the correct conclusion. Students of the PB were better at stoichiometric calculations, so it is likely that these students may have mainly worked with calculations rather than qualitative reasoning. This assumption is supported by the fact that there was a large drop in the mean score in 3.c compared to 3.a and 3b for the PB: 37.9 versus 60.6 and 47.1%. In contrast, in the ThB, the mean score in 3.c was about the same with 3.b: 22.8 versus 20.9%. In addition, in the TB the score in 3.c (22.0%) was similar to that in 3.a (24.2%) and much higher than in 3b: (10.2%).

Finally, in Part 4 we observe low mean scores: 31.9% (in 4.a), 34.4 (in 4.b), 21.5 (in 4.c). These questions involve demanding stoichiometric computations that required knowledge of organic reactions. The lowest score of 4.c might be caused by a misconception generated by the textbook (Kapetanou & Mavropoulos, 1998) that the students used. This book stated that “if a brown-red solution of bromine in carbon tetrachloride is added to an unsaturated compound containing a double bond, then the bromine solution is decolorized”. Of course,

this statement assumes that stoichiometric amounts of the compounds would react, or that an excess of the unsaturated compound would be added. But using the data provided with the problem, the correct answer is that the bromine solution is not decolorized because it was present in excess!

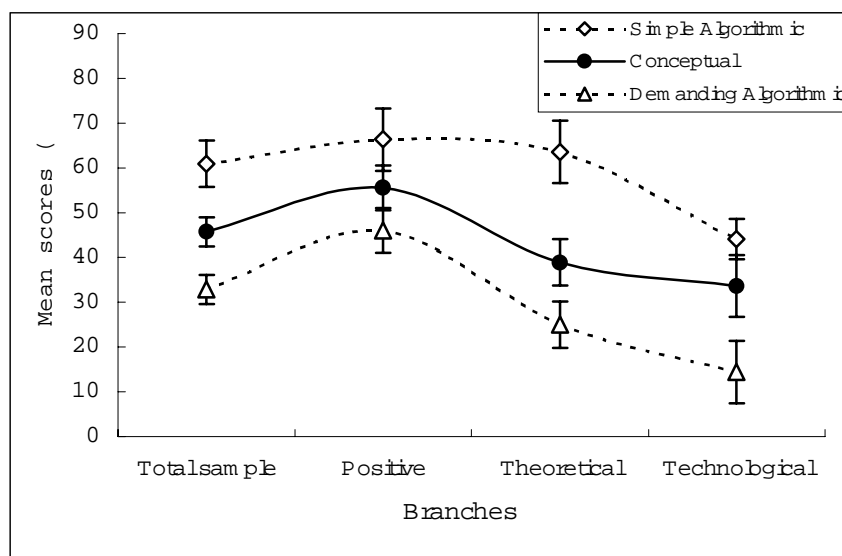
In all the questions studied here (Table 1), the PB had higher achievement than the ThB, and the ThB higher than the TB. Statistical comparison of the three branches by means of one-way ANOVA for independent samples and use of Tukey and Schaffe's tests confirms that in all questions (except 2.3.b where the difference between PB and ThB is not significant), the PB had statistically significant superior performance at $p = 0.01$ to both the ThB and the TB. On the other hand, the ThB had significantly superior performance to the TB in questions 3.b, 4.a, 4.b, and 4.c (at $p = 0.05$).

Conceptual versus Algorithmic

Figure 1 shows percentage mean scores of the total sample as well as of the separate branches in the sets of conceptual (2.3.a and 2.3.b), demanding algorithmic (3.a., 3.b, 4.a, 4.b, and 4.c) and simple algorithmic (1.6-1.8, 2.1 and 2.2) questions. The simple algorithmic questions were easy and produced the highest scores. On the other hand, in the case of the demanding algorithmic questions we had an escalation of complexity, so the achievement here was lower than in the two conceptual questions. The latter questions were obviously not very demanding. This sets, of course, a limitation to this study: had there been more demanding conceptual questions, the expectation is that the achievement in them would have dropped. In the relevant literature there is ample evidence of such a reversal. Note that in the case of the PB, the difference in the mean scores in the two kinds of questions is significantly smaller. This is due to the fact that in this branch we had much higher proportion of students with both abilities high (see below – Table 3).

The two bipolar dimensions, *Conceptual* and *Algorithmic*, were used to describe subject achievement. The students were divided into four categories based on their conceptual understanding and their algorithmic problem solving ability, according to Nakhleh (1993). Criteria for categorization were their performance on the questions that required conceptual

Figure 1. Percentage mean scores of the total sample as well as of the separate branches in the sets of conceptual, simple algorithmic, and demanding algorithmic questions. Error bars are included.



understanding (2.3.a and 2.3b) and those demanding algorithmic problem solving ability (3.a, 3.b, 4.a, 4.b, and 4.c). In this way, each individual was placed into one of four categories (see Table 2): High Algorithmic/High Conceptual (*AICI*), High Algorithmic/Low Conceptual (*AIC0*), Low Algorithmic/High Conceptual (*A0CI*), and Low Algorithmic/Low Conceptual (*A0C0*).

In order to assign each subject to the proper bipolar category, the following statistical criteria were used: average score (*M*), standard deviation (*SD*) and confidence limits (*CL*) at significance level $p=0.05$ were calculated. All subjects with scores greater than $M + CL$ were considered 'High', and all subjects with scores lower than $M - CL$ were considered 'Low'. In this way, a zone of subjects with scores $M - CL \leq \text{score} \leq M + CL$ was excluded; these subjects are grouped in a fifth separate category with the code *A*C**. The categorization and the frequency distributions are shown in Table 2. It is observed that that about half (52.1%) of the subjects demonstrated low conceptual thinking abilities (*C0*) and about half also (55.1%) demonstrated low algorithmic abilities (*A0*), while 43.1% were weak in both these abilities (*A0B0*).

Table 2. The categorization and the frequency distribution of Algorithmic versus Conceptual abilities.

<i>AICI</i> : 24.7 %	<i>AIC0</i> : 9.0 %	<i>AI</i>: 33.7 %
<i>A0CI</i> : 12.0%	<i>A0C0</i> : 43.1%	<i>A0</i>: 55.1 %
<i>CI</i>: 36.7 %	<i>C0</i>: 52.1 %	Totals: 88.8 %

Note that the missing 11.2 % belongs to the *A*C** category that was excluded from the categorization.

Of great interest is the distribution of the five groups amongst the three branches. Table 3 has the relevant data, together with the scores. As expected, the PB had by far the highest proportion of students high in both abilities. In addition, all three branches were about equivalent in the share of students who were high in conceptual ability only, while the PB and the ThB had about the same proportion of students high in algorithmic ability only. The PB had the highest number of students with algorithmic ability (irrespective of the conceptual ability): 45.7% versus 27.3% for the ThB and 15.1% for the TB. Similarly, the PB had also much higher number of students with conceptual ability (irrespective of the algorithmic ability): 47.9% versus 27.9% and 24.8%.

Next, we compare the mean total scores of the five categories among the three branches. One-way ANOVA for independent samples and use of Tukey and Schaffe's tests shows that in most cases the differences are statistically significant. In almost all cases, the PB had higher scores than the ThB, and the ThB higher than the TB. These differences could be attributed on the one hand to different characteristics of the students who opted for one branch or another, and on the other to the different educational experiences of the students in the three branches that made them more or less prepared for the examinations. This issue is addressed in more detail in the Concluding Comments.

Finally, we compare the corresponding distribution and achievement of the students of the PB in the two examinations (organic chemistry for general education and special examination in general chemistry), taking into account that we have different samples of the PB in the two cases. With regard to the distribution in the Nakhleh table (Table 3 in this paper, and Table 5 in Stamovlasis et al., 2004), we observe a reversal of the figures resulting from the two examinations. This can be attributed to the special features of the organic chemistry paper as

Table 3. Frequency distribution among categories and branches, and corresponding mean scores in the whole examination paper.

	Positive Branch*		Theoretical Branch		Technological Branch	
	Frequency (%)	<i>M</i> (<i>SD</i>)	Frequency (%)	<i>M</i> (<i>SD</i>)	Frequency (%)	<i>M</i> (<i>SD</i>)
<i>AICI</i>	35.9	86.9 (11.3)	17.4	78.5 (14.4)	9.7	74.0 (8.4)
<i>AICO</i>	9.8	69.4 (11.5)	9.9	66.9 (14.4)	5.4	59.8 (17.4)
<i>AOCI</i>	12.0	49.5 (12.9)	10.5	41.6 (11.9)	15.1	39.6 (15.1)
<i>AOCO</i>	30.8	28.9 (11.4)	49.4	28.0 (10.9)	62.4	24.3 (8.6)
<i>A*C*</i>	11.5	53.1 (20.9)	12.8	44.0 (18.9)	7.5	47.7 (10.2)

* For the purpose of comparison, the corresponding mean scores of the students of the Positive Branch in the advanced examination was: *AICI*: $M = 91.0$, $SD = 8.9$; *AICO*: $M = 67.8$, $SD = 11.8$; *AOCI*: $M = 70.2$, $SD = 15.0$; *AOCO*: $M = 34.6$, $SD = 15.4$; *A*CI**: $M = 71.3$, $SD = 15.9$ (see Table 6 in Stamovlasis *et al.*, 2004).

stated in the Introduction. The data for comparison of the achievements are given in Table 3 and as a footnote to it (see also Table 6 in Stamovlasis *et al.*, 2004). We observe that the achievement of the *AICI*, *AICO*, and *AOCO* students were similar in the two examinations: 86.9 in organic chemistry versus 91.0 in general chemistry for *AICI*; 69.4 versus 67.8 for *AICO*; 28.9 versus 34.6 for *AOCO*. In contrast, the *AOCI* and the *A*C** students scored much higher in the general chemistry examination (70.2 versus 49.5 for the *AOCI*; 71.3 versus 53.1 for the *A*C**). Note that, using the *t*-statistic, with the exception of the *AICO* case, all other differences are statistically significant. Our results seem to reinforce Nakhleh's categorization of students. On the other hand, it is important to note that, although the view of Mason, Shell, and Crawley (1997) that it appears to be rare or unusual that a student who understands the concepts lacks the ability to deal with an algorithmic problem may be potentially correct, in practice we do encounter such students; they are the ones with a lack of interest and practice in mathematical manipulations.

A two-way MANOVA for the Conceptual and Algorithmic Dimensions

An experimental design with interaction was carried out to investigate the main effects and the possible interactions between the two variables (a two-way Multivariate Analysis of Variance). Table 4 shows that in all cases the main effects are statistically significant, and in most cases they explain a considerable portion of the variance. On the other hand, the interactions are statistically significant in the case of the most demanding Parts 3 and 4 of the test, but explain only a small portion of variance. In addition, we carried out a three-way ANOVA with the Conceptual and Algorithmic Dimensions and the three branches as independent variables and the total score as dependent variable. The interaction Algorithmic Dimension \times Branch had a $p = 0.161$, explaining 1.4% of the variance, while for the interaction Conceptual Dimension \times Branch $p = 0.361$ and variance explained = 0.9%. These results show no interaction among the branches and the *Conceptual/Algorithmic* Dimensions with regard to the total score. This is an overall characteristic, while statistically significant differences among the branches do exist (see relevant discussion of the data in Table 3).

Table 4. Main effects and interactions of *Conceptual* and *Algorithmic* Dimensions in a *two-way* MANOVA for the four parts and the total examination paper.

		<i>Algorithmic</i>	<i>Conceptual</i>	
		Main effects	Main effects	Interaction
<i>Total</i>	<i>p</i>	0.000	0.000	0.216NS
	% Variance Explained	75.7	25.4	1.2
<i>Part 1</i>	<i>p</i>	0.000	0.020	0.659 NS
	% Variance Explained	25.1	1.6	0.5
<i>Part 2</i>	<i>p</i>	0.000	0.000	0.621NS
	% Variance Explained	35.3	27.4	0.5
<i>Part 3</i>	<i>p</i>	0.000	0.000	0.01
	% Variance Explained	64.8	20.7	2.7
<i>Part 4</i>	<i>p</i>	0.000	0.003	0.000
	% Variance Explained	63.1	2.4	4.1

Concluding comments and recommendations

1. *Was the postulated categorization of questions (conceptual versus algorithmic) supported by the data?* In line with the findings of our previous study (Stamovlasis et al., 2004), the classification of the questions that came out from principal component analysis (PCA) (see Appendix 2) is in agreement with the classification that was based on their nature (knowledge recall, simple algorithmic, demanding algorithmic, or conceptual understanding) as identified by the researchers. The statistical treatment with PCA gave a marginal structure and might be due to the small number of the conceptual questions and their low difficulty. The above conclusion was further reinforced by multivariate analysis of variance (MANOVA), which showed small interactions between the conceptual and the algorithmic dimensions for the most demanding Parts 3 and 4 of the test.
2. *Distribution of conceptual thinkers in the various branches.* The Positive Branch (PB) had the highest number of students with algorithmic ability (irrespective of the conceptual ability), as well as a much higher number of students with conceptual ability (irrespective of the algorithmic ability). An important finding was that all three branches were about equal in the share of students who were high only in conceptual ability. Finally, the PB and the Theoretical Branch (ThB) had about the same proportion of students high in algorithmic ability.
3. *Achievement in the three branches.* As expected, the PB had the highest mean achievement, intermediate was the ThB and lowest the Technological Branch (TB).
4. *The operation of Nakhleh's Scheme.* The model of student categorization suggested by Nakhleh (1993) was found to be operating. It is interesting to note that there is a considerable number of students from the ThB (27.3%) who demonstrated high conceptual abilities. These may be the corresponding second-tier students mentioned by Nakhleh. It is then the instructor's responsibility to make chemistry more interesting and attract these students to science. In addition, a three-way ANOVA showed no significant interaction between the branches and the *conceptual/algorithmic* dimensions.
5. *Is competence in algorithmic problem solving connected with competence in conceptual problem solving (and vice versa)?* In agreement with our previous study (Stamovlasis et

al., 2004), the statistical analysis for comparison and interaction between the conceptual and the algorithmic questions of the test supported the independence between the conceptual dimension and algorithmic dimension. We could conclude then that competence in algorithmic problem solving may be independent of competence in conceptual questions. The interpretation of the statistical analysis is not that the two abilities cannot coexist in the same person, but that the level of performance in one dimension does not depend on the level of performance in the other dimension. Put another way, one dimension does not presuppose the other, that is, the algorithmic problem-solving ability does not presuppose conceptual understanding, and vice versa.

Of particular importance is to examine the cause(s) for the differences among the three branches:

- (i) Did the differences reflect the characteristics of the students who opted for one branch or another? or
- (ii) Were there different educational experiences for the students of the three branches that made them more or less prepared for the examinations?

Surely, the students of the different branches had different characteristics. For instance, those of the ThB on the average liked the humanities more, but did not like mathematics or science; recall also that the students of the TB were, on average, weaker students. On the other hand, the students of the three branches had different educational experiences: they followed a number of common courses, plus some special courses depending on the branch. Regarding chemistry, all students had a common (easier) chemistry course (“Chemistry for General Education”), but students of the PB had an additional (more advanced/harder) chemistry course (“Chemistry for the Positive Branch”). Though the contents of the two courses did not overlap (one dealt with organic chemistry, the other with general/physical chemistry), students of the PB were surely more attracted to and more experienced in chemistry, so that it is very likely that they had developed a ‘chemical-type thinking’ to a greater degree. (They had no more experience in organic chemistry at the stage under consideration). Finally, although all students had received training in dealing with knowledge questions and algorithmic exercises within and outside the school, it is true that students in the PB had paid more attention to chemistry (and possibly be given more attention by instructors) than the students in the other two branches.

In conclusion, this study has provided further evidence about the extent of differences between conceptual understanding and algorithmic problem solving, being in agreement and reinforcing the findings of our previous similar study (Stamovlasis *et al.*, 2004). While it was found that a considerable number of students did lack one or both of these abilities, it was encouraging to find that about a quarter of our sample demonstrated both. At this point, however, the limitations of the present study (in addition to the different student characteristics and educational experiences in the three branches discussed above) should be re-emphasized. While the algorithmic questions of the test, dealing with stoichiometric calculations in organic chemistry, were numerous and had a graded difficulty, including some demanding problems/exercises, the conceptual questions of the test were limited and not very demanding. It is very likely that with more demanding conceptual questions, the proportion of students who could deal with them would have shown further decline. In any case, and although our investigation was necessarily contextually and locally bound, the results of the national examination have provided further evidence supporting the distinction and different nature of algorithmic and conceptual questions.

Turning to the implications of this work, these cannot be different from those of similar previous work (see Stamovlasis *et al.*, 2004). Taking into account that lack of understanding makes conceptual questions difficult for most students, teachers and schoolbook authors

should place emphasis on providing students with an understanding of chemistry (Gillespie, 1997). In addition, all students, but especially those experiencing difficulty with conceptual questions, must continually be given practice, encouragement, and support for dealing with such questions, with the aims both to improve their capabilities and develop their confidence. Finally, combined HOCS and LOCS-type, formal and informal, examinations and tests are needed for challenging and fostering students to develop their HOCS capacity (Zoller, 1993). A proper balance of the two types of questions should be included; otherwise, students may skip the demanding items (if there are few) or have a massive failure and disappointment (if there are too many).

Note: At present, the structure of the upper secondary school system in Greece (that is, the distinction into the three branches or streams, Theoretical, Positive, and Technological) remains, but there has been a dramatic shift of students from the Positive to the Technological Branch, because the latter, while leading to the same higher-education institutions as the Positive Branch (except bio- and medical subjects), does not include chemistry and biology among its four special subjects, but two 'softer' courses, one on computer science and another on business administration. Note however, that since the time of the examination analyzed in this paper, many changes have been introduced in the number of examinations and course requirements.

References

- Anderson T.W. (1984), *An introduction to multivariate statistical analysis*, 2nd ed., Wiley, New York.
- Catell R.B. (1966), The scree test for the number of factors, *Multivariate Behavioral Research*, **1**, 245-276.
- Gillespie R. J. (1997), Commentary: Reforming the general chemistry textbook, *Journal of Chemical Education*, **74**, 484-485.
- Kapetanou E. and Mavropoulos A. (1998), *Chemistry for general education at 11th grade* (in Greek), Pedagogic Institute / OEDB, Athens.
- Kaiser H.F. (1958), The Varimax criterion for analytic rotation in factor analysis, *Psychometrika*, **23**, 187-200.
- Mason S.D., Shell D.F. and Crawley F.E. (1997), Differences in problem solving by nonscience majors in introductory chemistry on paired algorithmic-conceptual problems, *Journal of Research in Science Teaching*, **34**, 905-923.
- Nakhleh M.B. (1993), Are our students conceptual thinkers or algorithmic problem solvers?, *Journal of Chemical Education*, **70**, 52-55.
- National Education Standards (1993), *National Education Standards: Observe, interact, change, learn*, National Academic Press, Washington, DC.
- Niaz M. (1995), Relationship between student performance on conceptual and computational problems of chemical equilibrium, *International Journal of Science Education*, **17**, 343-355.
- Stamovlasis D., Tsaparlis G., Kamilatos C., Papaoikonomou D. and Zarotiadou E. (2004), Conceptual understanding versus algorithmic problem solving: A principal component analysis of a national examination, *The Chemical Educator*, **9**, 398-405.
- Tsaparlis G and Zoller U. (2003), Evaluation of higher vs. lower-order cognitive skills-type examinations in chemistry: implications for university in-class assessment and examinations, *University Chemistry Education*, **7**, 50-57.
- Zoller U. (1993), Lecture and learning: are they compatible? maybe for LOCS; unlikely for HOCS, *Journal of Chemical Education*, **70**, 195-197.
- Zoller U., Lubezky A., Nakhleh M.B., Tessier B. and Dori Y.J. (1995), Success on algorithmic and LOCS vs. conceptual chemistry exam questions, *Journal of Chemical Education*, **72**, 987-989.
- Zoller U. and Tsaparlis G. (1997), Higher- and lower-order cognitive skills: the case of chemistry, *Research in Science Education*, **27**, 117-130.

**Appendix 1: The organic chemistry paper for eleventh-grade
National Greek Examination (June 1999)**

PART 1 (25 marks)

Q-1.1. The general formula C_nH_{2n} ($n \geq 2$) applies to: (a) all non-cyclic hydrocarbons; (b) alkanes; (c) alkenes; (d) alkynes. (3 marks)

Q-1.2. Organic compound $CH_3CH(OH)CH_3$ is named: (a) propanol; (b) methylethylether; (c) propanal; (d) 2-propanol. (3 marks)

Q-1.3. The products of complete burning of ethanol are: (a) CO_2 , O_2 , and H_2 ; (b) CO_2 and H_2O ; (c) CO and H_2O ; (d) C , CO , CO_2 , and H_2O . (3 marks)

Q-1.4. Benzene is: (a) a saturated hydrocarbon; (b) an aromatic hydrocarbon; (c) an unsaturated hydrocarbon with two double bonds; (d) an unsaturated hydrocarbon with one triple bond. (3 marks)

Q-1.5. (a) Which phenomenon is called isomerism? (b) List by name the kinds of structural isomerism. (3 marks)

Q-1.6. Esterification is called the reaction between and to form and The reverse reaction to esterification is named (3 marks)

Q-1.7. Of the carbonyl compounds, are oxidized by Tollen's reagent (ammoniacal solution of silver nitrate), forming a mirror, while their isomeric do not react. (3 marks)

Q-1.8. For each monomeric species in (I) there is a correspondent polymeric species in (II). Write down these correspondences. [Each time, write down the capital letter for each species in (I), and next to it write down the corresponding lower-case letter in (II).] (4 marks)

(I) A: $CH_2=CH_2$; B: $HC\equiv CH$; C: $CH_2=CHCl$; D: $CH_2=CHCH_3$.

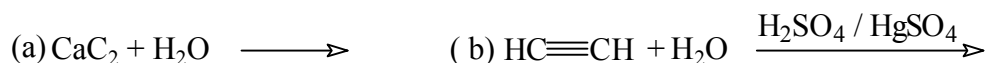
(II) a: PVC; b: bakelite; c: polyethylene; d: polypropylene; e: benzene.

PART 2 (25 marks)

Q-2.1. Explain whether or not the following statements are correct or wrong (9 marks):

- The molecular formula of a compound supplies more information than its structural formula.
- There are three organic compounds that correspond to the formula C_3H_8O .
- The compound with structural formula $CH_2=CH-CH_2CH_2-OH$ is named 1-buten-4-ol according to the IUPAC rules.

Q-2.2. Fill-in the following chemical equations for the reactions (8 marks):



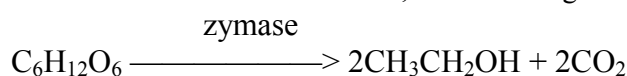
Q-2.3. On checking the exhaust gases of two cars A and B, it was found that the exhaust gases of A contain: CO_2 , water vapour, CO , hydrocarbons (C_8H_{18}) and nitrogen oxides, while the exhaust gases of B contain CO_2 , water vapour, and N_2 only.

2.3.a. Write the chemical equation for the reaction that explains the absence of hydrocarbons in the exhaust gases of car B. What catalyst is required for this reaction? (5 marks)

2.3.b. In which case (car A or car B) is it possible that volatile lead compounds may be detected? Explain. (3 marks)

PART 3 (25 marks)

A vessel contains an amount of must, which undergoes wine fermentation:



After completion of the fermentation, 200 L of wine 11.5° (11.5% v/v) were produced. The density of ethanol is $d = 0.8 \text{ g / mL}$.

- 3.a.** Calculate the volume and the mass of the alcohol produced. (10 marks)
3.b. Calculate the mass of the sugar ($C_6H_{12}O_6$) that underwent fermentation. (10 marks)
3.c. If you compared the (initial) mass of must with the mass of produced wine, would you find any difference? Explain. (5 marks)

Atomic weights: C : 12, H : 1, O : 16.

PART 4 (25 marks)

1.2 mol bromoethane (C_2H_5Br) is divided into three equal parts.

4.a. The first part of the bromoethane is dissolved to anhydrous ether, and an excess of sodium is added to the solution, producing an alkane (A). Calculate the mass of the produced alkane (A). (8 marks)

4.b. The second part of the bromoethane reacts completely with AgOH, and organic compound (B) is produced. Then excess of sodium reacts with B, when a new organic compound (C) is produced and a gas (D) is released. Calculate the volume of gas D under standard temperature and pressure (stp). (8 marks)

4.c. The third part of bromoethane reacts with excess of alcoholic (ethanoate) solution of KOH, and a gaseous hydrocarbon is produced. This gaseous hydrocarbon is then bubbled into a 1 L solution of bromine in tetrachloromethane (Br_2/CCl_4) 8% w/v. Will the bromine solution be decolorized? (9 marks)

Atomic weights: C : 12, H : 1, O : 16, Br : 80.

Appendix 2: A Principal Component Analysis of the questions

The data was treated with *Principal Component Analysis* (PCA) to reduce the number of variables and to detect structure in the relationships between variables, that is to classify variables (Anderson, 1984). We arrived at such a classification by looking at the correlation between variables and the factors (or 'new' variables) as they are extracted from the analysis; these correlations are also called factor *loadings*. Variables highly correlated with different factors are classified into different classes or categories. This pattern is referred as a *simple structure*.

In the literature, there are two criteria concerning the number of factors to be retained. According to Kaiser's criterion (Kaiser, 1958), one should retain only factors with eigenvalues greater than 1. The *scree test*, proposed by Catell (Catell, 1966), is a graphical method, using the plot of the factor eigenvalues (Figure 2). The aim is to find the place where the smooth decrease of eigenvalues appears to level out at the right of the plot. In addition the *Varimax* rotation method was used, which leads to a pattern of loadings on each factor that is as diverse as possible, leading itself to easier interpretation.

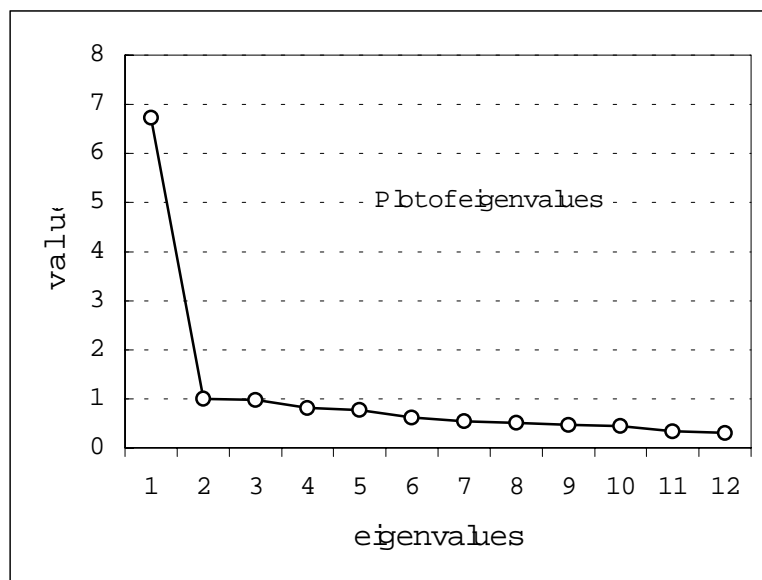
Despite the fact that two of the first three eigenvalues are marginally below unity, we find it useful to maintain the three factors. Note that the three factors explain significant portion of the variance. In addition the values of the *Kaiser-Meyer-Olkin Measure of Sampling Adequacy* (KMO Test) and *Bartlett's Test of Sphericity* (see also Stamovlasis *et al.*, 2004) were 0.958 and <0.001 (at least) respectively (using the SPSS 10.1 software). A value of the KMO test close to unity supports the usefulness of PCA; on the other hand, the value of Bartlett's Test is the significance level for rejecting the hypothesis that the initial variables are not related. Thus we are justified for using PCA as an analytical tool. Table 5 gives the results of the PCA for the whole sample with total explained variance of 62.2%. The classification of the questions that came out from PCA is in agreement with classification based on their nature as judged by the researchers (knowledge recall, simple algorithmic, demanding algorithmic, or conceptual understanding).

Factor 3 loads on knowledge-recall questions 1.1-1.4 and 1.5 that required just rote learning. This is also shown by the fact that in these questions we had the highest mean scores: 93.3% in 1.1-1.4 and 85.2% in 1.5.

Factor 1 loads on the questions that demanded applying of algorithms: 1.6-1.8, 2.1, 2.2, 3a, 3b, 4a, 4b, 4c. Questions 1.6 and 1.7 demanded knowledge of organic reactions, and thus achievement was moderate ($M = 56.1\%$). Question 1.8 involved correspondence between monomers and polymers ($M = 74.5$). Question 2.1 was about molecular formulas, of the right-wrong type with explanation, so achievement was moderate ($M = 56.2\%$); similarly question 2.2 referred to two organic reactions, and led also to moderate achievement ($M = 51.7\%$). The chemical equation that enters question 2.3.a justifies its partial algorithmic character and the considerable loading (0.41) by (algorithmic) Factor 1, hence its much lower loading by (conceptual) Factor 2 (0.60) compared to 2.3.b (0.80).

Finally, *Factor 2* loads mainly on conceptual questions 2.3.a and 2.3.b, and partially to 3.c. Question 3.c. does not belong to a single factor but is divided between factors 1 and 2; this is explained by taking into account that 3.c could be treated both as a conceptual and an algorithmic question (see above).

The classification of the questions of interest to this work was not altered on extracting just two factors, which show a clear separation between the algorithmic and the conceptual questions.

Figure 2. Plot of eigenvalues used for the scree test.**Table 5.** Principal Component analysis. Total sample ($N = 499$).

	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>
<i>Eigenvalues</i>	6.728	0.995	0.981
% Variance explained	48.1	7.1	7.0
% Cumulative variance	48.1	55.2	62.2
	Factor loadings*		
<i>Question</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>
1.1-1.4	0.21	-0.03	0.79
1.5	0.11	0.27	0.67
1.6-1.7	0.64	0.24	0.38
1.8	0.71	-0.13	0.17
2.1	0.62	0.35	0.19
2.2	0.70	0.28	0.32
2.3.a	0.41	0.60	0.13
2.3.b	0.07	0.80	0.17
3.a	0.68	0.42	0.20
3.b	0.71	0.45	0.08
3.c	0.30	0.54	0.02
4.a	0.78	0.24	0.11
4.b	0.82	0.30	0.11
4.c	0.77	0.31	0.07

* Varimax normalized; Factor loadings ≥ 0.60 are shown in bold.