

HIGHER AND LOWER-ORDER THINKING SKILLS: THE CASE OF CHEMISTRY REVISITED

Georgios Tsaparis

Introduction

According to John Gilbert (Gilbert, 2006, pp. 957–958), all over the world, high content loads lead to (chemistry) curricula that “are too often aggregations of isolated facts” so that students do not know how to form connections between them and “acquire a sense of how to give meaning to what they are learning”. In addition, students “fail to solve problems using the same concepts when presented in different ways”, and at the same time lack “a sense of why they should learn the required material”. In line with the “Next Generation Science Standards” for the “Framework for K-12 Science Education” (Krajcik et al., 2014; National Research Council [NRC], 2012), the emphasis should be on the in-depth study of a smaller number of disciplinary core ideas, instead of a mere transfer of discipline knowledge and the coverage of a large body of facts. The aims should be, on the one hand the explanation of physical phenomena, and on the other hand the practice in the methods of scientists.

The instructional methodology that is used by the teachers is crucial for providing students with the opportunity to develop abilities to deal with cognitively demanding problems and tasks, so as to contribute to the development in students of the so-called higher-order thinking (HOT)/higher-order thinking skills (HOTS) (alternatively termed by others *higher-order cognitive skills* (HOCS). The intellectual work in which the students are engaged will need to be different from what is going on in the science classrooms today (National Research Council, [NRC], 2015).

An integral part of the teaching-learning process is student assessment and examinations. According to Zoller and Tsaparis (1997), two major trends should drive reform of science education: (1) the belief that it is vital for students to develop HOCS; (2) that students should construct a deep conceptual understanding of the science they study rather than simply learning to apply algorithms to problem sets (Rutherford and Ahlgren, 1990; Zoller et al., 1995). HOT and HOTS should be contrasted to LOT and LOTS (Lower-Order Thinking/Lower-Order Thinking Skills) [Lower-Order Cognitive Skills (LOCS) for Zoller]. Examinations and other assessment tools should then be consonant with the above teaching/instructional goals, but also, meaningfully, contribute towards their attainment (Tsaparis & Zoller, 2003). According to Jensen et al. (2014), students who are tested throughout the semester with questions requiring HOTS, acquire not only deep conceptual understanding of the material, but also better memory for the course information.



JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Abstract. *This work analyses students' failure in the 2019 Nationwide Chemistry Examination in Greece, which concerns secondary education graduates, competing for admission to higher education Greek institutions. The distinction of thinking skills into higher and lower order (HOTS and LOTS) is used as a theoretical tool for this analysis. The examination included several questions that contained HOTS elements that had been unusual in previous examinations. This led to a decrease in overall student performance but better discrimination between outstanding and good students. Based on two samples of examination papers, corresponding to very similar subsets of the student population, the 2018 and 2019 examinations are compared, and the individual 2019 questions are evaluated. It was found that section B of the 2019 examination paper (which included contexts unfamiliar to the students, and for which, a large effect size between 2018 and 2019 was calculated) may have caused the large drop. An important link is established between the 2019 low performance and the HOTS and LOTS features of the questions, and the role or non-role of algorithmic calculations is examined. In addition, the critical opinions of chemistry teachers are provided, with a consensus emerging in favour of connecting chemistry with everyday life.*

Keywords: *chemistry examinations, higher-order cognitive skills, higher-order thinking skills, student assessment, twelfth-grade chemistry.*

Georgios Tsaparis
University of Ioannina, Greece



The following operational definition of LOCS and HOCS examination questions is borrowed from the Zoller and Tsapralis (1997) study:

LOCS questions: Knowledge questions that require simple recall of information or simple application of known theory or knowledge to familiar situations and contexts; they can also include problems (mostly computational exercises), solvable by means of algorithmic processes [mechanistic application of learned/recalled/known but not necessarily understood procedures (algorithms)] already familiar to the student through previous specific directives or practice or both (Bodner, 1987).

HOCS questions: Quantitative problems or qualitative conceptual questions, unfamiliar to the student, that require more than knowledge and application of known algorithms; they require analysis, synthesis, and problem solving capabilities, making of connections and critical evaluative thinking (Zoller et al., 1995), including the application of known theory or knowledge to unfamiliar situations or situations with an unusual element or dimension (DeCaprariis, 1978). Such an application may further require (partially or fully) the abilities of reasoning, decision-making, and critical thinking.

Bloom's initial taxonomy in the cognitive domain (Bloom & Krathwohl, 1956) was a basic tool in the Zoller and Tsapralis (1997) study for the above definition of LOCS and HOCS questions (henceforth to be referred to as LOTS and HOTS in this paper). This taxonomy distinguished the following dimensions: 'knowledge', 'comprehension', 'application', 'analysis', 'synthesis', and 'evaluation'. Anderson et al. (2001) revised the initial Bloom taxonomy, by replacing 'knowledge' with 'remembering' and 'synthesis' with 'creating', and by placing 'evaluating' before 'creating'. The various levels are not independent of each other, so there is an overlap between them. Note however that our definition of HOTS and LOTS questions goes beyond the Bloom taxonomy, with HOTS including *unfamiliar contexts* and also required to (deeply) *understand* the complex areas of chemistry (*conceptual understanding*). It follows from the above assumptions that HOTS go beyond the stage of knowledge and understanding of the factual, conceptual, or procedural dimension of science into a multitude of other activities (Zohar & Dori, 2003). Critical thinking and the metacognitive dimension are also very important (Anderson et al., 2001; Schneider, 2004).

Purpose and Rationale

The purpose of this study is, first, to analyse the students' failure in the 2019 NCE in Greece, with emphasis on the distinction of thinking skills into higher and lower order. Secondly, to compare the 2018 and 2019 NCEs and to evaluate the individual 2019 NCE questions, on the basis of samples of student scripts (examination papers). In addition, the 2019 NCE questions will be classified according to the requirement of HOTS or LOTS, and patterns in the data will be explored. Finally, the opinions of experienced teachers about the NCE will be reported.

Although prior research (including our own – hence the "Revisited" in the title of this paper) has established the connection of HOTS and LOTS with student performance in examinations, this study provides a remarkable correlation between the order of thinking skills (higher or lower) and student performance, so it reinforces previous findings. Further, even if it might appear that many aspects of the context are specific to the exact curriculum and examination approach in Greece, one should take into account that the questions cover standard general chemistry content. In addition, they are highly novel and interesting, providing support that the findings offer an explanatory, as well as a predictive tool for the difficulty of questions and students' performance in them. They can, therefore, inform instructional practice and assessment design, and in this way be of potential relevance and utility to a broad international community of future exam designers, practicing teachers, and textbook authors.

The 2019 Nationwide Chemistry Examination in Greece: The Most Difficult of the Last Decade?

In 2019, a fundamental change occurred in the Greek Nationwide Chemistry Examination (NCE), with a considerable number of questions involving the use of entities (such as aspirin, yogurt, and the ^{18}O isotope), in a way which was unusual in the previous examinations, introducing features of HOTS questions, such as the requirement of higher levels of the Bloom taxonomy (analysing and evaluating) plus the *application* of known theory or knowledge to *unfamiliar situations* or *situations with an unusual element or dimension* (DeCaprariis, 1978). See the Appendix for a selective sample of questions.

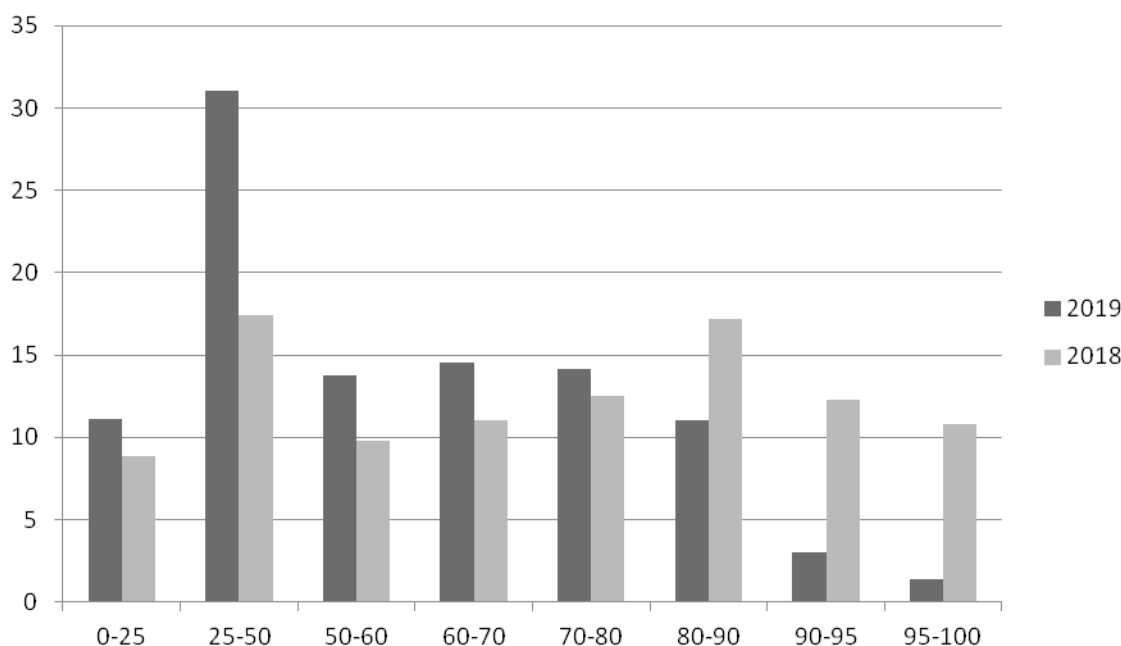
As a result, a large drop in student performance occurred, which caused an outcry on the part of the chemistry teachers. Here are some of the comments that were made on the internet: "A debacle of marks happened with chemistry – there must be a drop in the number of excellent performers" – "This (chemistry) is the subject with biggest downward difference in performance compared to last year: 42.12% (of the students) wrote below the 50% base level, while last year the rate was 26.25%" – "The questions for chemistry were particularly demanding, perhaps the most difficult of the last decade" – "The chemistry questions were a tough nut".



The bar graph of Figure 1 provides a breakdown of the chemistry marks for the 'Positive Stream' of studies for the years 2019 and 2018. (The "Positive Stream" concerns students who are candidates for medical, engineering, science, and agro-science higher education Greek institutions.) One should note first that 12,791 out of 30,368 students (42.12%) achieved less than the base mark of 50% in 2019. The corresponding figure for 2018 was 26.25%, $N = 31632$. At the other end, the very low percentage of students who achieved high marks in 2019 is remarkable: 2.99% for $90\% \leq \text{mark} < 95$ and 1.38% for $95\% \leq \text{mark} \leq 100$. The corresponding figures for 2018 were 12.28 and 10.82%. The comparison is striking against the 2019 examination, which scored far more poor marks and far less good marks, but, at the same time, had a more symmetrical distribution of the marks.

Figure 1

Data and comparison of performances for the 2018 and the 2019 chemistry examinations



Note: Student percentages are on the vertical axis, while mark percentage ranges are on the horizontal axis.

Source of data: Greek Ministry of Education and Religions (in Greek).

<https://www.minedu.gov.gr/rss/42070-28-06-19-anakoinosi-statistikon-stoixeion-gia-tis-vathmologikes-epidoseis-gel-kai-epal-2019>

The Questions on the 2019 Chemistry Examination

By tradition, the NCE consists of four 'sections'. Each section (say section B) consists of a number of questions (e.g. section B of the 2019 NCE consisted of questions B1, B2, B3, and B4). Many of the questions are further divided into parts (e.g. question B4 is divided into parts B4a and B4b). In this paper, as a rule, the general term 'question' is used to denote both a 'question' and a 'part' of a question. A selective sample of questions from the 2019 NCE paper is shown in the *Appendix*. The time duration of the examination was three (3) hours.

The Federation of Private Tutor Teachers of Greece (OEFE), in its official website, rated the 2019 NCE questions as "particularly demanding, perhaps the most difficult of the last decade - students need to have an in-depth understanding of the details of theory. (They were) particularly complex questions both in terms of their scientific content and the (required) time management".

The Association of Greek Chemists commented on the questions under the heading "Good intentions are not enough". Highlights: "The questions are the most difficult in recent years, displaying discrimination only among top students - they do not require memorization, they cover all the subject matter and require high critical ability". Other points: "It is positive to connect with everyday life and the applications of chemistry ... (but) it is not fair



that there is a discrepancy in the way we teach at school and the examination – Some topics have ambiguities, notably of scientific interest, which could cause problems and delays to the highly prepared students – The questions caused a ‘time suffocation’, even for excellent students – Some questions approached the philosophy of the Pan-Hellenic Student Chemistry Competition (PSCC), departing from that of the Nationwide Examinations, ... – The main problem lies with the fact that the (Greek) school cannot prepare the students so that to cope with similar questions. ... Some questions appear marginally outside the curriculum”.

As stated above, many questions in the 2019 NCE had features of HOTS (such as making of comparisons, analysing, and evaluating). In addition, many questions concerned contextual knowledge, unfamiliar to the students, which was esoteric to the questions, and influenced or could affect the given and goals of the respective questions. This was unusual in previous examinations, where the context was decorative, not affecting the given and the goals of the questions, so the chemistry could be answered without understanding (or even reading) that part of the question.

The tentative finding of the present study that section B of the 2019 chemistry examination paper was the major source of the reduction in marks in 2019 should be taken into account. For instance, question B4b, which referred to the ^{18}O ($^*\text{O}$) isotope, and asked “in which substance or substances of a mixture at (chemical) equilibrium will the $^*\text{O}$ isotope be detected” (with justification of the response), proved to be of particular difficulty. Apart from the ‘involvement’ of the concept of an isotope, which seems to have been unfamiliar to the students (it is mentioned in chapter 5 of the 10th-grade Greek high school textbook, but in recent years this chapter had been excluded from instruction), there is also the question of the *dynamic nature of chemical equilibrium*. The difficulty in understanding the sub-microscopic level of chemistry, the transition from the macroscopic to the sub-microscopic level, and the connection of the two levels are well known from the literature to cause many difficulties to students.

Research Questions

The following were the main research questions:

1. What causal factor(s) could be identified to explain the difficulty of section B?
2. Is student performance on the questions in the 2019 examination related to the HOTS or LOTS nature of the questions?
3. Are teachers’ opinions in line with the approach and the findings of the present study or are their views exaggerated?

Research Methodology

Expert Teacher-Assessors’ Opinions about the 2019 Nationwide Chemistry Examination - Method

In order to have a first-hand approach to the NCEs, four experienced teacher-chemists were consulted, who teach high school chemistry and have served as markers of students’ examination papers for several years. These teachers are referred to below as [T1], [T2], [T3] and [T4]. All four serve in public high schools in major Greek cities. Table 1 provides further information about the teachers. The teachers replied to a questionnaire written by the researcher, and the questions that are relevant to this paper are listed in the box below:

Table 1
Information about the participant teachers

	Gender	Years teaching	Bachelor’s degree	Postgraduate degrees
[T1]	Female	12.5	Chemistry	Master’s ChemEd, PhD ChemEd
[T2]	Male	24	Chemistry	–
[T3]	Male	30	Chemistry	Master’s ChemEd
[T4]	Female	36	Chemistry	–

Box: The teacher questionnaire

- (1) Comment on the data in Figure 1.
- (2) The 2019 questions have been described as, «perhaps the most difficult of the last decade». Do you agree with this view? If yes, for what reason? If no, for what reason?
- (3) List 2019 questions that you find particularly demanding compared to previous years. For what reason?
- (4) It has been suggested that the 2019 questions were "particularly demanding regarding time-management". Do you agree with this? If yes, for what reason? If no, for what reason?
- (5) List 2019 questions that you find particularly demanding in terms of time management.
You can add at the end any other positive and/or negative comment about the 2019 questions.

The most important findings from the responses of the four teachers to the questionnaire are presented at various points in this article. In addition, comments made on the internet by various teachers immediately after the 2019 examination are reported. Of particular interest was the fact that one of the four teachers has the habit of keeping a personal record about the rated student scripts, with detailed data for the marks per question. In particular, for the purpose of this study, this teacher made available the records for the 2017, 2018 and 2019 NCEs. The analysis of these data is done separately below. The statistical calculations were made using the SPSS statistical package.

Ethical Considerations

The four teachers who answered the questionnaire were known to the author from previous collaborations and did so after an invitation from the author. They were informed about the research nature and the purpose of this investigation, and agreed orally that their contribution is acknowledged in this paper with their names mentioned, although within the text codes are used, so that it is not revealed who said what. In the paper, selected parts of statements and opinions of two Greek public bodies as well of four Greek educationists are reported, whose names are mentioned here because the full statements and opinions are available in the public domain (the internet). In the paper, links to these are provided. The questions of the 2019 chemistry examination paper as well as the statistical data for the Greek national examinations are available in the public domain (the internet). The names of the 56 students whose exam scripts were used in this work were not available to the teacher who marked them (along with a second teacher) and consequently to the author of this paper, while the teacher who provided the performance data for these students did so on his own initiative and agreed that this data could be used by the author for the purpose of this work. Finally, the name of a foreign colleague who saw and commented on the 2019 NCE paper is reported here along with the comment and agreed in writing about this.

The 2019 NCE Questions and the Questions of the Nationwide Student Chemistry Competition

Several teachers compared (some of) the questions of the 2019 NCE with those of the Pan-Hellenic Student Chemistry Competition (PSCC) in Greece. Both examinations are held annually across Greece but have quite different features: NCE is targeted at secondary education graduates, who compete for admission to higher education in Greek institutions, while PSCC is aimed at the selection of the students who will represent Greece in the International Chemistry Olympics. The Zoller and Tsaparlis study (1997) used PSCC to evaluate the NCEs, for the year 1991 (for the PSCC: $N = 1352$). In the above study, all NCE questions were considered as LOTS questions, while the questions of the PSCC were distinguished into HOTS and LOTS: 22 (58%) HOTS questions and 16 (42%) LOTS questions. Focusing on the 146 students who scored at least the 50% mark level in the PSCC, it was found that their performance on the LOTS questions was much higher (+17.9%) than on the HOTS questions. In addition, no correlation was found (Spearman's coefficient $\rho = -0.01$) between students' scores on the HOTS and LOTS questions. The conclusion was reached that the HOTS and LOTS questions of the PSCC measured completely different skills, namely HOTS versus LOTS. This was also confirmed by a statistical factor analysis, which extracted



two factors, one factor loading on the overall PSCC score and the HOTS questions, and a second factor loading on the NCE, the overall PSCC score, and the LOTS questions.

Finally, it is useful and appropriate to refer to another study of ours (Stamovlasis et al., 2004), which analysed the NCE results from a conceptual understanding versus algorithmic problem solving perspective (Nakhleh, 1993). Our sample consisted of 647 high-school students in the “*Positive Stream*” of studies in Greece. Principal component analysis was used as a statistical tool to distinguish between the two perspectives. Performances were similar in the case of the conceptual and the more demanding algorithmic questions. On the other hand, easy recall and simple knowledge application questions were separated from all other questions. In subsequent work, a similar analysis was performed with a sample of 499 high school students from all streams of study (“Positive”, “Technological”, and “Theoretical”) (Stamovlasis et al., 2005).

Research Results

Data Analysis of Two Student Samples from the 2018 and 2019 Nationwide Chemistry Examinations

The first step is to analyse the detailed data from the student examination scripts marked by one teacher. First, note that the scripts that are marked in a given marking centre come from another region of the country, and the composition of the schools of origin is unknown. It should also be noted that the same scripts are marked by a second marker and that in case of a difference in the two marks for a script of 15% or more, a third marker remarks the script. In our case, none of the scripts required a third marking. For the three years 2017, 2018 and 2019, the particular teacher marked 50, 75 and 56 scripts respectively, for which a score in the range of 75-100% was obtained by 35.00, 52.00 and 35.71 percent of the scripts respectively. Comparison with national data (48.20 / 46.69 / 22.31% for the three years respectively) shows that for 2017 our sample consisted of below the corresponding national average mark, while for 2018 and 2019 it consisted of above the corresponding national average mark. For this reason, the 2017 sample will not be dealt with further, while the samples for 2018 and 2019 are of particular interest.

Comparison of Performance for the 2018 and 2019 Examinations

Table 2 shows the average percentages per ‘section’ (with standard deviations in parentheses) for the two samples from the 2018 and 2019 NCEs. It is noted that the examinations have a similar structure each year (and with their difficulty escalating from the easiest section A to most difficult section D), so the performances per section are directly comparable. It is observed that the mean scores in the two years for sections A, C and D are similar and do not show statistically significant differences as judged by means of the Mann-Whitney U test:¹ for A(2019) v. A(2018), $p = .453$; for C(2019) v. C(2018), $p = .897$; for D(2019) v. D(2018), $p = .936$. Further, if one takes the averages of sections A, C and D, one sees that they have a remarkable coincidence (71.6 vs. 71.1 / standard deviation 22.2 vs. 22.7; $p = .920$). This shows that the scripts marked by the teacher in 2018 and 2019 correspond to very similar subsets of the student population, although, notably with higher performance than the national population. A large and statistically significant difference is found in section B: for B(2019) v. B(2018), $U = 968$, $z = 5.264$, $p < .001$. The η^2 (eta squared) value is .212, accounting for a large *effect size* (Lenhard & Lenhard, 2016).

The important conclusion is that, based on the specific data available to us, the problem of a large drop in performance in the 2019 examination was caused by section B. Of course, relying on two random, but not necessarily representative samples, with 75 and 56 students respectively, constitutes a significant limitation, which leads us to draw the above conclusion with caution.

1 Because the scores in all cases (except for the total score for 2019) did not follow the normal distribution (as judged by means of the Shapiro-Wilk test), the Mann-Whitney nonparametric test was used. The statistical conclusions are the same as when using the parametric *t* test for independent samples.



Table 2

Performance data (averages with standard deviations in parentheses) for the student samples from the 2018 and 2019 Nationwide Chemistry Examinations. All numbers are percentages

2019 (N = 56)						
Section	A	B	C	D	A+C+D	A+B+C+D
	86.1	39.1	66.3	60.9	71.1	63.1
	(18.7)	(27.9)	(30.8)	(27.9)	(22.7)	(22.7)
2018 (N = 75)						
Section	A	B	C	D	A+C+D	A+B+C+D
	87.7	67.1	68.2	58.9	71.6	70.5
	(19.4)	(25.2)	(26.3)	(31.6)	(22.2)	(22.9)
Comparison 2019 to 2018: Mann-Whitney test: z-statistic value for independent samples and p-values (two-tailed test)						
z	.754	5.264	-.128	-.0768	.195	1.845
p	.453	<.001	.897	.936	.920	.066
Comparison 2019 to 2018: statistical t-values for independent samples and p-values (two-tailed, with unequal variances assumed)						
t	-0.496	-5.907	-.367	.379	-.131	-1.710
p	.621	<.001	.714	.705	.896	.090

Note. All sections A, B, C and D are equivalent, with each contributing 25% to the overall score.

Finally, it is also worth comparing the performances on the individual sections A, B, C and D for 2019. Using the non-parametric Friedman test for repeated measures, a statistically significant difference is found: $p < .001$. Further, comparisons between pairs of sections for 2019, based on the non-parametric Wilcoxon signed ranks test, lead to statistically significant differences for all pairs ($p < .001$, except for the C-D pair for which $p < .05$): $A > B$, $A > C$, $A > D$, $B < C$, $B < D$, $C > D$. It is evident that, for 2019, section B disrupts the gradual lowering of marks, being much more difficult than the others. The data for 2018 shows that, with the exception of the B-C pair where an almost equivalent performance occurs ($p = .508$), questions got harder as one moves from Section A to section D.

Performance on the Individual Questions of the 2019 Examination

Table 3 contains the findings for the detailed performance (averages with standard deviations in parentheses) of the student sample ($N = 56$) in the 2019 NCE. All marks are in percentages.

As expected, in section A (a set of five simple, multiple-choice, knowledge recall questions with no requirement to explain/justify the choices made), a very high performance happened, with the highest value (98.2%) for question A5 (it asks for the identification of the electronic structure of the ground state for the ^{18}O atom), while in question A2 (71.4%) students appeared to find difficulty in distinguishing between the connection or no connection of reaction enthalpy and of activation energy to the reaction rate.



Table 3

Analytical performance data (averages with standard deviations in parentheses) of the student sample (N = 56) from the 2019 exam. All figures are percentages

A1 (5)	A2(5)	A3(5)	A4(5)	A5(5)	Total A (25)
85.7 (35.3)	71.4 (45.6)	87.5 (33.4)	87.5 (33.4)	98.2 (13.4)	86.1 (18.7)
B1(5)	B2(6)	B3(6)	B4(8)		Total B (25)
42.9 (32.7)	38.1 (36.9)	42.8 (37.0)	34.4 (28.5)		39.1 (27.9)
C1 (13)	C2(5)	C3(7)			Total C (25)
67.4 (30.2)	59.3 (38.9)	69.1 (35.9)			66.3 (30.8)
D1 (3)	D2(6)	D3(7)	D4(2)	D5(7)	Total D (25)
78.6 (25.8)	49.1 (34.4)	59.4 (36.8)	70.5 (44.5)	62.2 (41.0)	60.9 (28.1)
Grand total (100)					63.1 (22.7)

Note. All sections A, B, C and D were equivalent, with each contributing 25% to the overall score. The questions were each assigned points, which are written in parentheses next to the number of each question. For example, question A1 was assigned 5 points, B2: 6 points, C1: 13 points, and so on.

Relatively good performances were found for questions D1 (about balancing two chemical equations and identifying the oxidizing and reducing agents, 78.6%), D4 (about the qualitative prediction, using the Le Chatelier principle, of the effect of a change in pressure on the position of chemical equilibrium, 70.5%), C3 (involving stoichiometric calculations, 69.1) and C1 (concerning organic reactions of pheromone, 67.4%). In section C, the lowest performance (59.3%) was in question C2, which asked: (C2a) to calculate the pH of a sample of a yogurt at the endpoint of the titration with a standard solution of NaOH; (C2b) to calculate the % w/w content of lactic acid in the given yogurt sample.

The lowest performances were in section B, with lowest being for questions B4 (34.4%) and B2 (38.1%). Question B4 consisted of two parts, B4a and B4b. Question B4b – the one with the ^{18}O isotope – has already been discussed. Question B4a involves two equilibrium calculations, one starting with the forward reaction and the other starting with the backward reaction: $\text{PbO(s)} + \text{CO(g)} \rightleftharpoons \text{Pb(l)} + \text{CO}_2\text{(g)}$. Question B2 also consists of two parts: B2a is easy, especially because of showing the carboxylic group $-\text{COOH}$, so the low mark must be attributed to B2b. B2b asked for an explanation of the great difference between the 1st ionization energy of boron and the energy for the 2nd ionization of carbon.

Questions B1 (about whether aspirin is more easily absorbed, in its non-ionic form, in the stomach or in the small intestine) and B3 (about graphs representing the volume of oxygen, which is released during the catalytic decomposition of a hydrogen peroxide solution) followed in difficulty with a similar performance (42.9 and 42.8% respectively). B1 consists of two parts" part B1a asked to write an equation for the chemical ionization reaction for aspirin, and it was easy because it shows the carboxylic group as $-\text{COOH}$, so the low mark must be attributed to B1b.

Finally, from section D, most difficult was question D2 (49.1%), which required stoichiometric calculations involving three chemical equations (it asked for calculation of the degree of conversion of the reactant into product as a fractional number). Relatively difficult, with similar difficulty (59.4 and 62.2% respectively) were (i) question D3 [with two easy parts D3a and D3b, and a composite chemical equilibrium problem] and (ii) question D5 (a composite ionic equilibrium problem). Both questions D3 and D5 deal with 'problems', which had been extensively practiced by the students. Finally, as noted above, question D1 was easy (78.6%) and question D4 relatively easy (70.5%). It is noteworthy that if one excludes question D2 [where a low performance occurred (49.1%), similar to questions B1, B2 and B3], the average of the remaining questions in section D (D1, D3, D4 and D5) is 67.7%, which is very close to overall performance in section C (66.3%).



At this point, a consideration will be made of the answers of the four teachers with regard to the questions that they found particularly demanding. Section B as a whole caused great difficulty for students ([T1] and [T2]), where there was the greatest failure of students [T2]. In particular (with the exception of [T4] who found that the questions “were more demanding as a whole and relative to the insufficient time”), the other three teachers reported on question B4b: “The students were confused by the addition of solid” [T2]. Also, isotope theory is only found in the tenth grade, but it was exempted from instructional material ([T1] and [T3]). According to [T3], B4b also requires an in-depth understanding of the concept of chemical equilibrium: “the addition of solid Pb^*O does not affect the position of the chemical equilibrium, but the (albeit in small amount) isotope is transferred to all oxygen-containing substances (due to the dynamic nature of the chemical equilibrium)”. For question B1 (about aspirin), “the absorption of a substance is a concept that most students were not aware of, as opposed to the neutralization concept, which they had encountered many times” [T1], also “students did not understand the common-ion effect” [T2]. The above explain the overall low performance in section B (see Table 4).

The Views of the Teachers

The views of the four teachers were similar. For them, “the questions were not clear in their wording” [T1] / “The style of the questions is markedly different (more demanding) than in previous years. (The diversification had already begun in 2017.)” [T3] / “The (questions) were more demanding as a whole and with regard to the rather inadequate time” [T4]. In addition, all teachers agreed with the view that the 2019 questions were probably the most difficult of the past decade:

[T1]: “The questions have been characterized as the most difficult in recent years. Some questions had the philosophy of the Pan-Hellenic Student Chemistry Competition (PSCC). (On the other hand, they were) in the right direction, because through their connection with life they were calling on the students to process daily topics and through the answers to test their understanding. The problem was that the way the textbook is written, and the way teaching is clearly described in the teachers’ book does not follow the logic of the questions. The students were asked to respond to questions that they were not trained to deal with”.

[T2]: “This (the difficulty of the questions) was due to the way the questions were worded (they had more similarity to the PSCC questions) and required a greater understanding of the topics than in previous years”.

[T3]: “They (the questions) refer to substances of biological, pharmaceutical or industrial interest. This originality probably disturbed the examinees, since their wording differs from... the exercises of the (standard) textbook, but also of the other study aids. They combine the theory from various (book) chapters (and) require ... not just knowledge of the theory but an in-depth understanding of it. (Such are) questions: B1b, B2b, B3, B4b, C1, C3, D3a”.

[T4]: “They differed greatly in style both from the textbook material and from previous years’ questions, which I think took the students by surprise. (On the other hand), they needed substantial knowledge, as they should do”.

The Problem of Adequacy of Time

Several teachers appeared to believe that insufficient time may have been provided for the examination, but no real data are available to support the time issue. However, it seems reasonable to expect that where questions are presented in an unusual context, students will probably need to spend more time thinking about them, unless they quickly decided to abandon them.

Other (Positive and Negative) Comments and Suggestions by the Teachers

[T1]: “The questions must be oriented towards the applications of chemistry in our daily lives (*positive*) and be included in the textbook, so that students can prepare (for the exams) appropriately (*negative / suggestion*). They must be more clearly worded” (*negative*).

[T2]: “(There has been a) greater connection of chemistry with reality (*positive*). (There was a) large number of redox reactions compared to previous years (*negative*). (There was a) poor escalation of the difficulty of the questions. In my view, section B should be less difficult, while section D should be more difficult (*negative*)”.

[T3]: “The large number of questions and their degree of difficulty should be (and probably were) such that any marker who studied the subject thoroughly and had sufficient practice in theoretical questions and exercises should take approximately 160 minutes to answer the questions fully and satisfactorily. This will ensure that the candidates



are evenly distributed on the marks scale, which is desirable (positive / proposal). In question C1 the information is incomplete, and the wording of D4 is likely to have confused some (negative)".

[T4]: "The questions required a deeper understanding of the material rather than just typical knowledge (*very positive*). They corrected considerably the right asymmetry in the distribution curves of previous years' scores (*positive*). Some carelessness (was noted) regarding the scientific perspective (*negative*)".

Return to HOTS and LOTS Types of Questions

The above findings from the sample data of 56 students from the 2019 NCE are used to make a direct connection of the questions to HOTS and LOTS. In Table 4 the questions for which scores were available are classified against the two competencies. The classification was made independently by the author and one of the four teachers, based on the definitions of HOTS and LOTS that are used in this work. The agreement was 78.6%. The few discrepancies were resolved through discussion between the two. Note that the teacher was not aware of the data from the sample of 56 students and most of the discrepancies between the author and the teacher were resolved in favour of the teacher's disposition. Therefore, the categorization of the questions as HOTS and LOTS must be assumed independent of the 56 students' performance.

The requirement of one or more HOTS, including some of the higher levels of the Bloom taxonomy (e.g. making of connections, analysing and evaluating) was the first criterion, and this concerned most of these questions. Also, unfamiliar context was another feature for classifying a question as HOTS. Note also that there were several questions that involved composite calculations, and which would definitely require HOTS if they were novel to the students, but they were classified as LOTS because the students had been given extensive practice in similar problems during their (at least) year-long preparation for the exams.

Looking for patterns between competencies and performance, one can easily find that the lowest scores are for the questions that in whole or in part correspond to HOTS (B1, B2, B3, B4, C2, D2, and D3). Focusing on Section B, the problem seems to arise, because of the requirement for conceptual understanding, critical thinking, the higher levels of Bloom's taxonomy, but also because of unfamiliarity with the context being used. It is also remarkable that in terms of the distribution of the marking points, 36 percentage points correspond to HOTS questions and 64 percentage points to LOTS questions (HOTS / LOTS ratio = $1/1.78 \approx 1/2$).

In the table there is an additional column entitled "Calculations". In it, reference is made to whether each question contains mathematical (algebraic) calculations or the balancing of chemical equations. The calculations are divided into simple and composite ones. The classification was empirical and made independently by the author and one of the four teachers ([T2]). The discrepancies between them were resolved through separate discussions between the author and two additional teachers ([T1] and [T3]).

Table 4

Classification of the 2019 questions into HOTS and LOTS types

		HOTS	LOTS	
Section (mean mark)	Question (mean mark)	Questions (assigned points)		Calculations
A (86.1)	A1 (85.7)		A1 (5)	
	A2 (71.4)		A2 (5)	
	A3 (87.5)		A3 (5)	
	A4 (87.5)		A4 (5)	
	A5 (98.2)		A5 (5)	
B (39.1)	B1 (42.9)	B1b (4)	B1a (1)	
	B2 (38.1)	B2b (4)	B2a (2)	
	B3 (42.8)	B3 (6)		
	B4 (34.4)	B4a (4), B4b (4)		B4a: composite



		HOTS	LOTS	
Section (mean mark)	Question (mean mark)	Questions (assigned points)		Calculations
C (66.3)	C1 (67.4)	C1a (7), C1b (3), C1c (1), C1d (2)		C1b: composite C1d: composite or simple (see text)
	C2 (59.3) C2a (2) C2b (3)			C2a: composite C2b: composite
	C3 (69.1)	C3 (7)		C3: composite
D (60.9)	D1 (78.6)	D1a (1) D1b (1) D1c (1)		D1a & D1b: simple or compos- ite (see text)
	D2 (49.1)	D2 (6)		D2: composite
	D3 (59.4)	D3c (3)	D3a (2) D3b (2)	D3b: simple D3c: composite
	D4 (70.5)		D4 (2)	
	D5 (62.2)		D5 (7)	D5: composite

Notes. Column 5 notes if calculations are involved. Columns 1 and 2 give in parentheses the mean scores of the sample of 56 students, while columns 3 and 4 give in parentheses the percentage marking points for each question or part of a question.

Regarding the findings, one observes first that questions that contain only HOTS or both HOTS and LOTS parts had the lowest scores (34-59%), regardless of whether or not they include calculations. On the other hand, questions that are LOTS-only have scores >70 (with the exception of a 62 mark, where complex calculations were involved). Generally, where the questions are LOTS not involving calculations (A and D4) or calculations are simple (D1), the scores are high or relatively high. The case of LOTS questions with composite calculations (C1, C3, D5) corresponds to intermediate marks. Finally, it is worth noting that B4, which had the lowest score (34%), included two HOTS parts, with both being conceptually demanding, especially B4b.

At this point, it is necessary to return to the conclusion that, on the basis of the data from the 56 scripts, the sharp drop in performance in the 2019 NCE was caused by section B. If, in table 4, one counts the assigned points as they are divided between the HOTS and LOTS questions, it is seen that for section B there are 22 HOTS and only 3 LOTS points (HOTS: 22/25 = 88%). In contrast, LOTS outperforms HOTS in the other three sections: 100% for section A, 80% for section C, and 64% for section D. These results reinforce the conclusion about the increased difficulty of section B. In addition, they are in line with the differences in performance between sections A, C, and D.

Discussion

Reference should be made first to a recent work (in Greek) by Giannakopoulos et al. (2019). This work applied the SOLO taxonomy to the 2019 NCE questions, and also compared these with the questions for the 2017 and the 2018 examinations. It was found that most (18) of the questions for 2019 correspond to the combined relational and the extended abstract SOLO levels, while fewer questions (8) correspond to the combined uni-structural and multi-structural levels. The conclusion was that the questions for 2019 "required high levels of understanding, quality knowledge of the subject matter, and high critical ability" (pp. 9-10). In addition, the comparison of the questions for 2017, 2018, and 2019 showed "a tendency to shift from the uni-structural and multi-structural SOLO levels, mainly to the relational level, reaching as well the extended abstract level and with "a gradual increase" of the abstract questions, which were non-existent before 2017" (pp. 9-10).

In the present research, selective literature was used to show the important role that examinations can play in distinguishing between questions requiring HOTS or LOTS skills, with LOTS questions being expected to lead to higher student performance than HOTS questions. Several of the questions on the 2019 Greek Nationwide Chemistry Examinations (NCE) can be assumed to be of the HOTS type (by involving the application of known



theory or knowledge to unfamiliar contexts, and also by requiring conceptual understanding, critical thinking, and one or more of the higher-level Bloom dimensions (applying, analysing and evaluating), and this is expected to have contributed to the lower performance. The independent comments by an academic non-Greek colleague (Dr. Bill Byers) are also relevant: "The questions are interesting and demanding. There is an absence of simple information recall and understanding, and application of knowledge are needed. Both qualitative and quantitative reasoning is required. I would expect this examination paper to discriminate well between the stronger and weaker candidates". Connecting chemistry to everyday life must also have been uppermost in the minds of the examiners. Unfortunately, teachers and secondary school students were not informed of the intended change. On the other hand, the examination board did not comment on the results, nor has it suggested what might be expected on the 2020 paper.

On the basis of a sample of 56 student scripts, it was found that section B of the examination paper might have caused the large drop in student performance. Further, 2019 individual questions were analysed. Despite the necessity, non-representative, rather small sample, Tables 2, 3 and 4 contain interesting and original findings that shed light on the problem being discussed. The comparisons made are revealing and allow for certain conclusions. This answers our first study question about "*what causal factor(s) could be identified to explain the difficulty of section B?*"

An immediate and extremely important feature of the performances in the 2019 NCE emerged through the classification of the questions according to their HOTS and LOTS demands, and further examining the role or non-role of algorithmic calculations. The HOTS questions had the lowest scores regardless of whether they included calculations or not. Also, questions that were LOTS-only had the highest scores, while LOTS questions with composite calculations corresponded to intermediate scores. The problem with section B appears to be due not only to the requirement of various HOTS, but also, in many cases, to students' lack of familiarity with the context being used. Section B was assigned by far the highest percentage of HOTS marking points (88%), reinforcing the conclusion about its increased difficulty. It is also remarkable that question B4b (which definitely is a difficult one, requiring HOTS and the connection of the macro with the sub-micro levels of chemistry) can be considered a special case because "isotopes" might not even have been taught. The above conclusions provide the answer to our second research question ("*Is student performance on the questions in the 2019 examination related to the HOTS or LOTS nature of the questions?*").

At this point it should be emphasized that, although the previous discussion demonstrates that the HOTS and LOTS demands provide a useful approach to the analysis of the questions, there are limits to its application: first, there is a great variation of questions that can be classified as HOTS, and secondly there is a variation in degree to which a question can be classified as HOTS or LOTS, that is this categorization cannot be quantified. In addition, the HOTS-LOTS categorization is a function of the students' background and the sort of teaching they have been exposed to in class so a question that requires HOTS for some students may require LOTS to others in a different context. Thus, many of the questions of sections C and D would definitely require HOTS if they were novel to the students, but they were categorized as LOTS in Table 4 because of the extensive (at least year-long) student practice in such questions.

The limitation of the non-representative and rather small sample of the present study should also be reiterated. Further research is required that by using a large and highly representative sample of students in the same or similar future examinations worldwide could shed new light on various relevant issues, for instance the effect on student performance of ample practice in HOTS tasks throughout their education. Other possible prospective research could examine the role of various psychometric variables of students, such as working memory capacity, mental (*M*) capacity, Piaget's level of cognitive development (i.e., general hypothetical ability), degree of perceptual field dependence/independence ('disembedding ability') (Demerouti, et al., 2004; Tsapalis, 1998, 2005; Tsapalis & Angelopoulos, 2000; Tsapalis, Kousathana, et al., 1998).

Considering the third research question ("*Are teachers' opinions in line with the approach and the findings of the present study or are their views exaggerated?*"), a variety of views were identified, and they seem reasonable and thoughtful. There was a clear consensus that the questions in 2019 were the most difficult in recent times, and, in general, the teachers were right in their criticism. Indeed, there were ambiguities/incomplete wording, in some questions (and even an error in one question), but most people would agree with those who think that there must be a smooth distribution of the students on the scoring scale. Regarding the ambiguities and scientific errors, a general dilemma appears where a task/problem is solved as intended by the examiners, but the underlying science/reality is deficient or even flawed. It happens at times to come across such issues in examinations, so examiners should be encouraged to be as scientifically rigorous as possible.

The teachers pointed out (actually 'protested') that "the way the [standard] textbook is written, and the way teaching is described in the teachers' book does not follow the logic of the questions. The students were



asked to respond to questions that they were not trained to deal with". One should not expect, however, that students should be given exhaustive practice in such questions in an attempt and hope that what the students will actually encounter in the exams will not be HOTS but only LOTS questions. As prior education research has shown, "teachers often tend "to reduce high level demands on students' thinking by breaking a demanding task into less challenging subtasks, by proceduralizing ["algorithmisizing"] the work for the students, or by focusing on the correctness and completeness of the tasks rather than students' sense making" (Tekkumru-Kisa et al., 2018, p. 480, and references therein).

In the education literature there have been several studies that were concerned with the use of various instructional methods for promoting HOTS in students: Barnett and Francis (2012); Bramwell-Lalor and Rainford (2014); Ealy (2016), Eklund and Prat-Resina (2014), Flynn (2014), Ghani-et al. (2017), Lavery et al. (2016), Prat-Resina, 2020, Saito et al. (2018), Tekkumru-Kisa et al. (2018), Zoller and Pusking (2006). According to Tekkumru-Kisa et al. (2018, p. 480, and references therein), "the instructional tasks selected by teachers and assigned to students in science classrooms play a critical role in shaping students' opportunities to learn science as emphasized in the Framework [for K-12 Science Education" (National Research Council [NRC], 2012)] ... Different tasks will demand different kinds and levels of student thinking and engagement in the discipline". Further, "consistent results across more than a decade of classroom research suggest the need for students to engage with cognitively demanding tasks in order to develop the kind of rigorous student thinking promoted in the 'Framework'".

Returning to the Greek case, the author agrees (first and foremost - see below) with the view of Stratos Stratigakis, an education analyst of Greek nationwide examinations, that "the 2019 difficult chemistry questions were an entirely correct 'trick' in order to substantially drop the baseline and 'distinguish the excellent' students from the very good ones". One should not then openly condemn the innovations introduced in the 2019 Greek chemistry examination. Last but by no means least, it would be very interesting to consider how many of the 420 students who achieved a mark in the range 95-100% on the 2019 chemistry national examinations actually chose to enter departments of chemistry or even physics, mathematics, or biology? In fact, very few - the vast majority opted for medical and engineering schools!

Conclusions and Implications

It is evident that the binary division of the questions into HOTS and LOTS-types offers both an explanatory and a predictive tool for the difficulty of questions and students' performance on them. Three things stand out:

- 1) use of unfamiliar contexts may cause problems;
- 2) formative assessment should be appropriate to subsequent summative assessment;
- 3) the education process should contribute to developing HOTS in students.

In particular, this study has reinforced the conviction that students who depend heavily on algorithms are often unable to cope with even minor changes to the way in which questions are asked. There also appears to be a consensus that chemistry should be linked to real life issues. Finally, although the examination and the data that were used in the present study were set in a given national context, both their analysis and the findings are generalizable to other countries and provide useful information to education policy makers, examiners, examination boards and teachers that can guide them towards informed decisions and help future examining and teaching.

What are the implications of the findings for education policy and for instruction? Referring to John Gilbert's argument that students do not know how to "acquire a sense of how to give meaning to what they are learning", and at the same time "fail to solve problems using the same concepts when presented in different ways", one should be warned that, for many chemistry teachers, the mastery of computational, LOTS-type exercises (traditionally referred to as 'problems') is assumed as 'equivalent' to conceptual understanding of chemistry. However, conceptual understanding requires complex upper-class thinking. It must then be accepted that instruction and examinations should contribute to the development/acquisition of HOCS [HOTS] by our students at all levels.

Both HOTS and LOTS questions should be generally used for class evaluation, that is, this approach should apply not only to final/summative assessments, but also to the formative evaluations of learning. This should be done very early in the education process (starting with elementary school) and apply to all subjects: the essence is to promote the practice and development of HOTS in students. Obviously, much better planning of teaching and much better teaching methodology are required to promote the construction of deeper knowledge. Finally, the importance of linking chemistry to everyday life and applications, thus making the lessons more interesting for students should be stressed, and this should happen throughout the year, and not just appear on the final examinations., while books should be enriched with relevant topics and examples.



Note

Interested readers can receive a copy of "Supplementary Information" with all items of the 2019 examination paper, with comments (Part I), plus "Justifications and comments about the HOTS/LOTS classification of the questions" (Part II) by e-mailing the author.

Acknowledgments

The author is grateful to the four teacher-chemists *Aristides Kafkias, Elli Lampri, Eleni T. Pappa, and Basileios Zikovellis* for answering the questionnaire and for the very helpful discussions he had with them. (The alphabetical order of their names does not coincide with the name-codes order [T1], [T2], [T3], [T4].) Special additional thanks are due to *Aristides Kafkias*, who provided me with his marking data for the Greek Nationwide Chemistry Examinations for the years 2017, 2018 and 2019. Dr *Eleni Pappa* collaborated in the classification of the questions on the 2019 chemistry examination into HOTS and LOTS, and many thanks are due to her for this. Without their contribution, this work would have been limited to a theoretical and bibliographical approach. Finally, the author acknowledges the help of: Dr *Giannoula Pantazi* for executing the statistical computations; associate professor *Michael Siskos* for discussions about organic chemistry; and Dr Bill Byers for reading carefully the manuscript and commenting on it, and for many thoughtful and very useful discussions he had with him about this work.

References

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201-211. <https://www.tandfonline.com/doi/abs/10.1080/01443410.2011.638619>
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals*. Longman.
- Bodner, G. (1987). The role of algorithms in teaching problem solving. *Journal of Chemical Education*, 64(6), 513-514. <https://doi.org/10.1021/ed064p513>
- Bramwell-Lalor, S., & Rainford, M. (2014). The effects of using concept mapping for improving advanced level biology students' lower- and higher-order cognitive skills. *International Journal of Science Education*, 36(5), 839-864. <https://doi.org/10.1080/09500693.2013.829255>
- DeCaprariis, P. (1978). Identification of students' abilities. *The Physics Teacher*, 16(6) 387-388.
- Demerouti, M., Kousathana, M., & Tsaparlis, G. (2004). Acid-base equilibria, Part II. Effect of developmental level and disembedding ability on students' conceptual understanding and problem-solving ability. *The Chemical Educator*, 9(2) 132-137. <http://chemeducator.org/bibs/0009002/920132gt.htm>
- Ealy, J. B. (2016). Visualization of kinetics: Stimulating higher-order thinking via visualization. *Journal of Chemical Education*, 93(2), 394-396. <https://doi.org/10.1021/acs.jchemed.5b00215>
- Eklund, B., & Prat-Resina, X. (2014). ChemEd X Data: Exposing students to open scientific data for higher-order thinking and self-regulated learning. *Journal of Chemical Education*, 91(9) 1501-1504. <https://doi.org/10.1021/ed500316m>
- Flynn, A. B. (2014). How do students work through organic synthesis learning activities? *Chemistry Education Research and Practice*, 15(4), 747-762. <https://doi.org/10.1039/c4rp00143e>
- Ghani, I. B. A., Ibrahim, N. H., Yahaya, N. A., & Surif, J. (2017). Enhancing students' HOTS in laboratory educational activity by using concept map as an alternative assessment tool. *Chemistry Education Research and Practice*, 18(4), 849-874. <https://doi.org/10.1039/c7rp00120g>
- Giannakopoulos, N., Klavdianos, D., Sinigalias, P., & Spyroulias, G. (2019). Critical evaluation of the shift of the chemistry questions for the 2017 to 2019 Pan-Hellenic Examinations, *Chimica Chronica*, Jul-Aug 2019, 81(6) 8-12 (in Greek). <https://www.eex.gr/news/anakoinwseis/2317-teuxos-iouliou-augoustou-2019>
- Gilbert, J. K. (2006). On the nature of "context" in chemical education. *International Journal of Science Education*, 28(9) 957-976. <https://doi.org/10.1080/09500690600702470>
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test ... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307-329. <https://doi.org/10.1007/s10648-013-9248-9>
- Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, 25(2), 157-175. <https://link.springer.com/article/10.1007/s10972-014-9383-2>
- Lavery, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol, *PLoS ONE* 11(9): e0162333. <https://doi.org/10.1371/journal.pone.0162333>



- Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. https://www.psychometrica.de/effect_size.html
- Nakhleh, M. B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry. *Journal of Chemical Education*, 70(1) 52-55. <https://doi.org/10.1021/ed070p52>
- National Research Council [NRC] (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on conceptual framework for New K-12 science education standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press. <https://www.nap.edu/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts>
- National Research Council [NRC] (2015). *Guide to implementing the next generation science standards*. Committee on guidance on implementing the next generation science standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press. <https://doi.org/10.17226/18802>
- Niaz, M. (1995). Relationship between student performance on conceptual and computational problems of chemical equilibrium. *International Journal of Science Education*, 17(3) 343-355. <https://doi.org/10.1080/0950069950170306>
- Prat-Resina, X. (2020). Using data-driven activities with ChemEd X Data to practice structure-property relationships in general chemistry. *Chemistry Teacher International*, Ahead of Print. <https://doi.org/10.1515/cti-2018-0010>
- Rutherford, F. J., & Ahlgre, A. (1990). *Science for all Americans*. Oxford University Press. <http://www.project2061.org/publications/sfaa/online/sfaatoc.htm>
- Saido, G. A. M., Siraj, S., DeWitt, D., & Al-Amedy, O. S. (2018). Development of an instructional model for higher order thinking in science among secondary school students: A fuzzy Delphi approach. *International Journal of Science Education*, 40(8), 847-866. <https://doi.org/10.1080/09500693.2018.1452307>
- Schneider, W. (2004). Metacognitive development: Educational implications. *Encyclopedia of the Social & Behavioral Sciences*, 9730-9733. <https://doi.org/10.1016/B0-08-043076-7/02388-3>
- Stamovlasis, D., Tsapalis, G., Kamilatos, C., Papaoikonomou, D., & Zarotiadiou, E. (2004). Conceptual understanding versus algorithmic problem solving: A principal component analysis of a national examination. *The Chemical Educator*, 9(6) 398-405. <https://doi.org/10.1333/s00897040825a>
- Stamovlasis, D., Tsapalis, G., Kamilatos, C., Papaoikonomou, D., & Zarotiadiou, E. (2005). Conceptual understanding versus algorithmic problem solving: Further evidence from a national chemistry examination. *Chemistry Education Research and Practice*, 6(2) 104-118. <https://pubs.rsc.org/en/content/articlelanding/2005/rp/b2rp90001g#!divAbstract>
- Tekumru-Kisa, M., Stein, M. K., & Coker, R. (2018). Teachers' learning to facilitate high-level student thinking: Impact of a video-based professional development. *Journal of Research in Science Teaching*, 55(4), 479-502. <https://doi.org/10.1002/tea.21427>
- Tsapalis, G. (1998). Dimensional analysis and predictive models in problem solving. *International Journal of Science Education*, 20(3) 335-350. <https://doi.org/10.1080/0950069980200306>
- Tsapalis, G. (2005). Non-algorithmic quantitative problem solving in university physical chemistry: A correlation study of the role of selective cognitive factors. *Research and Science and Technological Education*, 23(2) 125-148. <https://doi.org/10.1080/02635140500266369>
- Tsapalis, G., & Angelopoulos, V. (2000). A model of problem-solving: its operation, validity, and usefulness in the case of organic-synthesis problems. *Science Education*, 84(2), 151-153. [https://doi.org/10.1002/\(SICI\)1098-237X\(200003\)84:2%3C131::AID-SCE1%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2%3C131::AID-SCE1%3E3.0.CO;2-4)
- Tsapalis, G., Kousathana, M., & Niaz, M. (1998). Molecular-equilibrium problems: Manipulation of logical structure and of M-demand, and their effect on student performance. *Science Education*, 82(4) 437-454. [https://doi.org/10.1002/\(SICI\)1098-237X\(199807\)82:4%3C437::AID-SCE2%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1098-237X(199807)82:4%3C437::AID-SCE2%3E3.0.CO;2-C)
- Tsapalis, G., & 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(2) 50-57.
- Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, 12(2), 145-181. https://doi.org/10.1207/S15327809JLS1202_1
- Zoller, U., & Pushkin, D. (2007). Matching higher-order cognitive skills (HOCS) promotion goals with problem-based laboratory practice in a freshman organic chemistry course. *Chemistry Education Research and Practice*, 8(2), 153-171. <https://pubs.rsc.org/en/content/articlelanding/2007/rp/b6rp90028c#!divAbstract>
- Zoller, U., Lubezky, A., Nakhleh, M. B., Tessier, B., & Dori, J. (1995). Success on algorithmic and LOCS vs. conceptual chemistry exam questions. *Journal of Chemical Education*, 72(11) 987-989. <https://doi.org/10.1021/ed072p987>
- Zoller, U., & Tsapalis, G. (1997). Higher- and lower-order cognitive skills: The case of chemistry. *Research in Science Education*, 27(1), 117-130. <https://doi.org/10.1007/BF02463036>



Appendix

Selective examples from the 2019 Chemistry Examination Paper

(Note. In the actual paper, the marks/points assigned to each question and part of question were shown – These are also shown in Table 4.)

SECTION A (Note. One question is shown here from a total of five.)

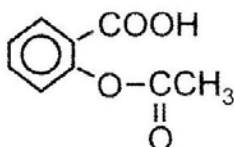
A3. Acidic bioactive substances may cause ulcers in the stomach. Which of the following substances is most likely to cause ulceration of the stomach?

- atorvastatin ($pK_a = 4.5$)
- oestradiol ($pK_a = 10.4$)
- paracetamol ($pK_a = 9.5$)
- phenobarbital ($pK_a = 7.4$)

SECTION B (Note. Three questions are shown here from a total of four.)

B1.

Aspirin

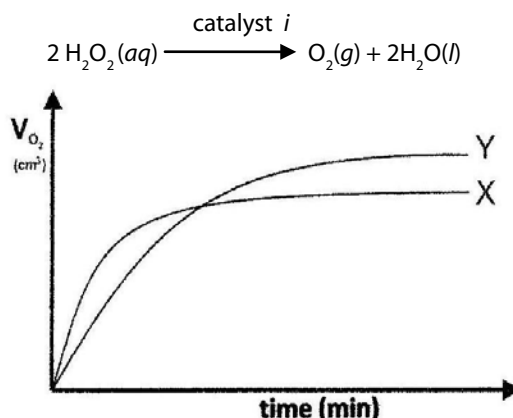


is a weak organic acid, which ionizes in the aqueous environment of the gastrointestinal tract.

B1a. Write an equation to show the chemical ionization reaction for aspirin.

B1b. Aspirin is more easily absorbed in its non-ionic form. Explain where it will be absorbed more: in the stomach, where the $pH = 1.5$, or in the small intestine, where the $pH = 8$.

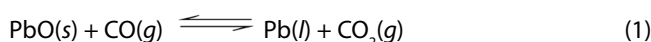
B3. Curve X of the following graph represents the volume of oxygen (O_2), which is released during the catalytic decomposition of a 1 M hydrogen peroxide solution as a function of time. The reaction is:



Explain which of the following changes would lead to the production of curve Y:

- Addition of H_2O .
- Addition of a 0.1M solution of H_2O_2 .
- Use of a different catalyst (catalyst ii).
- Decreasing the temperature.

B4. The following equilibrium is given:



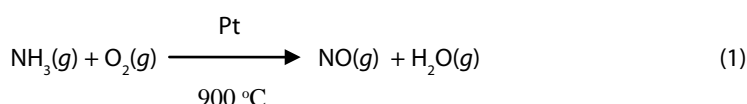
B4a. 1 mol of PbO(s) and 1 mol of CO(g) are introduced into a fixed volume vessel, and 1 mol of Pb(l) and 1 mol of CO₂(g) are introduced into a second vessel of the same volume. The two vessels are heated to a suitable temperature θ and equilibrium (1) is established. Compare the quantities of CO(g) in the two vessels. Justify your answer.

B4b. ¹⁸₈O is an isotope of ₈O. The isotope ¹⁸₈O can be denoted as *O. It is possible in the lab to know whether a molecule carries this isotope or not. In one of the above vessels (sub-question B4a), in which equilibrium (1) has been established, a small amount of Pb*O(s) is introduced. After the passage of some time, in which (one or more) substances of the equilibrium mixture will isotope *O be detected? Justify your answer.

(Note. **Section C** is not shown here – It consisted of three questions.)

Section D (Notes. Two main questions are shown here from a total of five – The information provided concerns also the other three questions.)

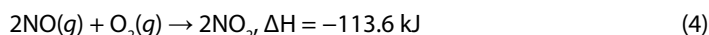
Nitric acid is a chemical compound of major importance to the global economy. It is mainly used (75% of world production) for the production of NH₄NO₃, a component of fertilizers. The modern method of manufacturing nitric acid is based on conversion of ammonia to nitric acid in a three-stage process. The first step involves catalytic oxidation of ammonia to nitrogen monoxide (Ostwald process):



Balance this reaction.

...

The second step of the process involves the oxidation of NO to NO₂ according to the equation:



...

D4. The third step of the method is as follows:



Explain whether the nitric acid synthesis reaction (5) is favoured at high or low pressure.

D5. After the reaction of NO₂ with H₂O, a 10M HNO₃ solution is obtained. If you have available a 5M aqueous NH₃ solution, calculate the volume ratio with which the two solutions must be mixed to arrive at a neutral solution. You are given that: All solutions are at the temperature of $\theta = 25\text{ }^\circ\text{C}$, $K_b(\text{NH}_3) = 10^{-5}$, $K_w = 10^{-14}$. Appropriate approximations can be used.

Received: February 27, 2020

Accepted: May 25, 2020

Cite as: Tsapalis, G. (2020). Higher and lower-order thinking skills: the case of chemistry revisited. *Journal of Baltic Science Education*, 19(3), 467-483. <https://doi.org/10.33225/jbse/20.19.467>

Georgios Tsapalis

M.Sc., PhD, Professor Emeritus of Science Education, Department of Chemistry, University of Ioannina, GR-451 10, Ioannina, Greece.

E-mail: gtseper@cc.uoi.gr

Website: <http://users.uoi.gr/gtseper/>

ORCID: <https://orcid.org/0000-0002-0856-747X>

