Exploring Novel Tools for Assessing High School Students' Meaningful Understanding of Organic Reactions

Theodoros Vachliotis, Katerina Salta, Petroula Vasiliou, and Chryssa Tzougraki*
Department of Chemistry, University of Athens, Panepistimiopolis, Zografou, 15771, Athens, Greece
**tzougraki@chem.uoa.gr*

The systemic approach to teaching and learning (SATL) teaching model has been developed during the past decade *(1−5)*. SATL techniques have been implemented and evaluated in many different knowledge domains in all educational grades, but the major teaching applications have been reported on chemistry topics in secondary and tertiary education *(2−4)*. The results reported from the evaluation of SATL techniques are significant as far as the improvement of students' academic achievements *(2, 3)*. Epistemologically, SATL can be considered as a hybrid approach that combines and uses features and ideas from systemics and constructivism adjusted in concept mapping procedures. SATL originators recognize as the basic goal of this approach the achievement of meaningful learning by students and suggest that this goal can be attained through the development of systems thinking, in a context of constructivist and systemic-oriented learning tasks (SATL techniques) *(1−3, 5)*. In this direction, they have proposed new types of assessment questions, systemic assessment questions (SAQs), aiming to a more effective evaluation of the systemic-oriented objectives in the SATL model *(6, 7)*.

Theoretical Basis

Systemics is a broad term, which takes into account the fact that there is a range of different systems approaches. Most of these approaches offer not only a theory, but also a way of thinking (systems thinking) and a methodology for dealing with systemic issues *(8)*. Systems thinking provides a means to conceptualize the interaction of the multiple components that make up complex systems *(9)*. It encompasses a wide variety of research and fields of study, including general systems theory, system dynamics, complexity science, and systemics *(10)*. The construct of systems thinking has been emphasized as an understanding of the patterns that connect various systems ideas, methods, theories, or models *(11)*. Although it is often confused with systems theory, which takes a holistic perspective, systems thinking takes a more balanced approach, focusing on the whole as well as the parts to form a more complete understanding of the system *(12)*.

Meaningful learning was described by Ausubel as the formulation of nonarbitrary relationships between ideas in the learners' mind *(13)*. In surface or rote learning, learners do not construct relationships between concepts nor integrate new concepts to their prior knowledge. Instead, they rely on memorizing and employ a surface approach to learning. In contrast, in deep or meaningful learning, learners deal with a learning task by attempting to form relationships between newly and previously learned concepts *(14)*. In a classroom setting, someone could suggest that the implementation of systems thinking oriented teaching−learning procedures could have the potential to foster a linear curriculum by taking a more holistic perspective of the subject matter, thus, helping students to learn more meaningfully.

Considering the enhancement, and consequently the assessment, of learners' meaningful understanding of a scientific knowledge domain as an important educational goal, we initially investigated whether specific SAQ forms are potentially effective tools for assessing students' meaningful understanding of organic reactions in 11th grade core chemistry courses in Greek high schools. More than one reason prompted us to perform this study. To begin with, the major part of the high school organic chemistry curriculum refers to the production and chemical properties of organic compounds, namely, to various types of organic reactions. Moreover, meaningful understanding of organic chemistry topics in secondary education has so far not been adequately studied. To our knowledge, there are only three reports in the literature related to this theme. The first study *(15)* explored correlations between first-year students' understanding of concepts obtained in high school and their performance in university exams. The examined concepts were related to organic chemistry. However, the investigation did not focus on the validation of the assessment tools used. The second study *(16)* was not based on empirical research. Instead, a specific concept mapping procedure was proposed for promoting students' deeper understanding of organic reactions. In the third study *(3)*, a preliminary evaluation of SATL techniques' influence on students' academic performance on organic chemistry topics was reported. Although various SAQ types were included in the instrumentation used, the focus was not on the validation of SAQs for assessing meaningful learning, which is the purpose of our investigation.

Taking all these under consideration, we designed and conducted a study trying to obtain preliminary evidence for answering the following research question:

> Are specific forms of SAQs valid and reliable tools for assessing students' meaningful understanding of organic reactions in the 11th grade in high school?

Concept Mapping for Assessing Meaningful Learning

Cognitive psychology has demonstrated that the way knowledge is structured in memory determines the ability to retain, recall, and use it to solve problems *(17)*. Rote learning occurs when the learner makes little or no effort to relate new information to existing knowledge or to novel situations.

10.1021/ed9000415 Published on Web 01/07/2011

In contrast, meaningful learning occurs when the learner interprets, relates, and incorporates new information with existing knowledge and applies the new information to solve novel problems *(18)*. Meaningful learning seems to be related with the achievement of higher-order cognitive skills. If someone wishes to foster and assess meaningful learning, they need to emphasize those cognitive processes that go beyond recall of knowledge. In addition, when transfer of knowledge occurs, it indicates meaningful learning *(19)*.

Science education researchers have identified concept mapping as an assessment strategy having potential to elicit students' meaningful understanding *(20)*. The various concept map assessment systems have been categorized on the basis of two criteria: (i) the task that students perform to demonstrate and record their knowledge; and (ii) the scoring system used to evaluate students' knowledge *(21)*. The task format can be constrained or open-ended, with various intermediate possibilities. Constrained tasks are tasks that restrict the students to a supplied list of concepts, or link words, or use a fill-in-the-blank approach. Open-ended tasks supply a small number of prompt concepts, and otherwise do not restrict how the map may be drawn. Intermediate tasks are those that specify a list of concepts to be used, but place little or no other restrictions on how the map can be drawn. The methods for scoring concept maps generally combine an interest in the content validity or accuracy of the content displayed in the map with an interest in the elaborateness of the map as measured by counting various map components, such as concepts or links.

Two major concerns have arisen regarding scoring methods for concept maps. The first is that traditional methods *(22, 23)* are time-consuming and require the input of an expert, either in terms of judging the validity and importance of map components, or in the construction of a criterion map. One way by which this concern has been addressed is the development of simplified map-scoring techniques *(24)*. Another alternative is to analyze maps in terms of their overall structure rather than in terms of a detailed analysis of concepts, links, and propositions *(25)*. The second issue is concerned with psychometric properties of concept maps. Their reliability and validity are integrally related to the concept map task and to the scoring system used *(23)*. Some important conclusions about reliability and validity measures are that they are better for proposition-based scores than for structural scores *(26, 27)* and better as raters become more experienced *(28)*. As for construct validity, it has often been suggested that concept maps measure different aspects of knowledge than traditional assessment techniques, which do not measure meaningful learning. Thus, it is not surprising that moderate or low correlations with conventional types of assessment have been found *(26, 29)*.

Systemic Assessment Questions

The central role in the SATL model has the two-dimensional spatial arrangement and representation of concepts under study and their interrelationships. The model proposes an arrangement of concepts through closed, interacting, and evolving conceptual systems ("concept clusters"), in which all interrelationships are made explicit to the learner using a concept map-like representation. Such a closed conceptual system representation is called a "systemic diagram" and is the basic teaching tool proposed in the SATL model context *(2, 3, 5)*. Systemic assessment questions include systemic diagrams, which are subsystems of the systemic diagrams used for the instruction of the corresponding topic. Such a subsystem represents a constellation of concepts, that is, a single loop of closely interrelated concepts. The systemic diagrams of SAQs take various geometric forms, such as triangular, quadrilateral, pentagonal, hexagonal, and so on, depending on the number of concepts that are incorporated in the diagram (Figure 1) *(6, 7)*.

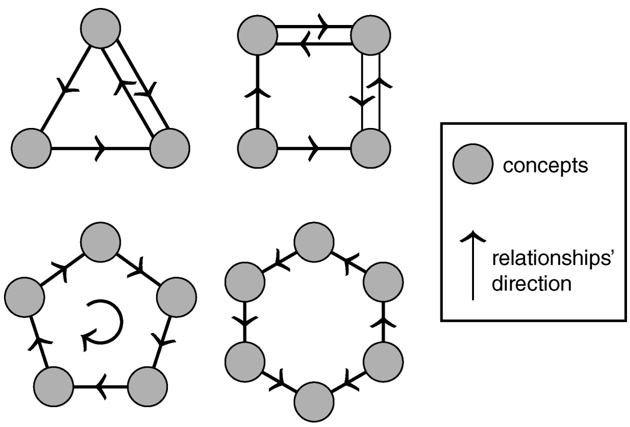

Figure 1. Geometric forms and symbols of systemic diagrams used in SAQs.

Taking a step forward from what it has been proposed, we suggest that the closed loop of concepts in some of the SAQ systemic diagrams could be extended, including concepts that are not part of this specific loop, but they are directly related with concepts belonging to this loop. Such a perspective facilitates the construction of high-quality systemic diagrams for almost every subject matter. Simultaneously, SAQs' basic systemic philosophy is better served, qualitatively as well as quantitatively. Of course, in such a case, the systemic diagrams in SAQs can also take other forms, besides the single-loop geometric forms described above. Examples showing how these suggestions can be applied in the construction of SAQs related to chemistry concepts' assessment are presented in the next section, where the two SAQs under study are illustrated.

SAQs definitely belong to the broader group of concept mapping techniques. According to the directedness continuum *(30)* of these techniques, which is related with *what is provided* to the examinees, most of the proposed SAQ types can be located closer to the high-directed edge of this continuum having a constrained task format, given that the map structure (spatial arrangement of concepts) is provided and only the completion with missing components or the analysis of the map to its components is required. Even for the SAQs in which the construction of a systemic diagram is required and only the included concepts are provided, the skeleton map is implied as there is a correspondence between the number of concepts and the polygonic structure of the diagram (*n* concepts correspond to a *n*-angle, single-loop geometric form). Moreover, in SAQ systemic diagrams, in contrast with classical concept maps, there are no hierarchical levels and consequently no cross-links either.

On the basis of the discussion above, the concern about classical concept maps and the assessment of these maps' structural characteristics are offset in SAQs. What is assessed is the knowledge of the concepts being involved in the diagram and their interrelationships, namely, the construction of valid propositions for the conceptual system under study. Furthermore, in SAQs only one predetermined right answer is accepted. Even for the "construct-a-diagram" type SAQs, in the required geometric loop only one specific arrangement of concepts under study is possible. As a result, SAQs' scoring can definitely be *objective*. These two characteristics of SAQs are very important because, in combination with a clear scoring method, they contribute to overcoming the major concerns about maps' scoring: the lack of practicality in scoring the maps (time-consuming, need for expertise) and the problems related with the psychometric properties of the maps.

Regardless of the fact that SAQs seem to offer the potential for overcoming these concerns, a fundamental issue needs to be investigated in order to use them for assessing meaningful learning: that is the matter of construct validity. The high directedness of most types of SAQs can be a disadvantage for reflecting differences among students' knowledge structures. A comparative study of two mapping techniques *(30)* concluded that these differences were better reflected by a "less-directed", "construct-a-map-from-scratch" technique, compared with a "more-directed", "fill-in-a-skeleton-map" technique. On the other hand, students using SAQs have to consider several concepts at once, apply the concepts in a new situation, and synthesize them creating a comprehensive answer, a meaningful conceptual whole. When formulating an answer to such a question, students not only must have developed an understanding of scientific concepts, they must also be able to select specific concepts that fit the particular test item and combine them into an integrated web of meaning, through the appropriate use of scientific language and symbols. With this reasoning, it seems that the answers to SAQs should illustrate students' meaningful understanding of a scientific topic. Moreover, in order to better determine the cognitive demands evoked by a mapping technique, it is important to consider all the characteristics of this technique: not only what is provided, but additionally the extent and amount of what is provided, the significance of what is provided, and what is required from the examinees *(31)*. Taking all these under consideration, we preliminarily investigated the effectiveness of two specific SAQ forms for capturing meaningful learning.

Methodology

Construction of Tests

In this study, we evaluated only the type of SAQs requiring the completion of semicompleted and structured systemic diagrams ("fill-in-the blanks" SAQs) with missing components not provided. This type of SAQ has a constrained task format and we selected it considering that it is more consistent with the conventional objective questions often used for assessment in high school, and thus more familiar to the students. Trying to answer our research question, we had to determine the characteristics of SAQs under study.

First, we had to select the number of concepts in SAQ systemic diagrams. While the number of concepts in these diagrams has no theoretical limit, psychologically, Miller's number 7 ± 2 *(32)* may well be a realistic limit for students. That is, the comfort zone in conceptualization of a system of concepts is about one-half dozen concepts. Conversely, given that the SAQ format under study is closer to the high-directed edge of the continuum, one could claim that one-half dozen concepts are not enough for capturing meaningful learning. Thus, we decided to comparatively evaluate two SAQs including systemic diagrams

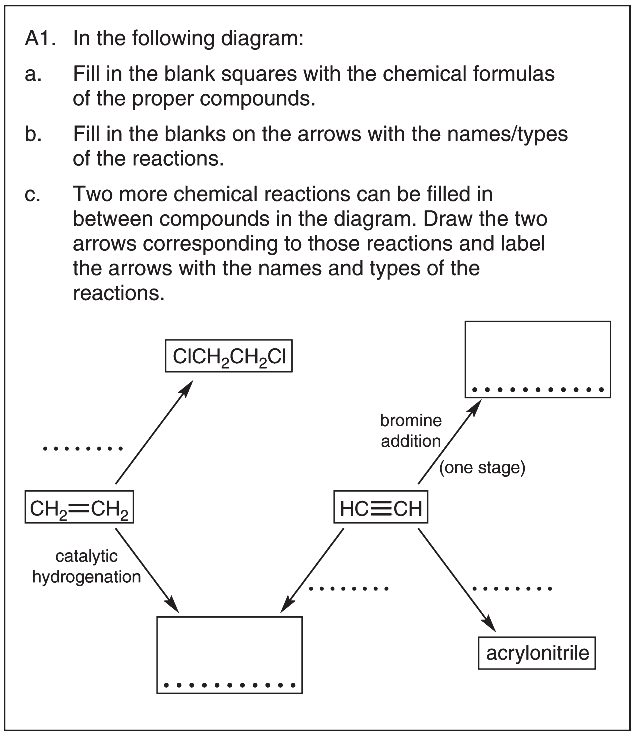

Figure 2. SAQ of the first (A) test.

with quite a different number of concepts: the first SAQ (Figure 2) consisting of 6 concepts (inside Miller's comfort zone), and the second SAQ (Figure 3) consisting of 11 concepts (above Miller's comfort zone). For this purpose, and given that data collection would be subsumed under the formal Greek high school two-semester assessment, we were oriented toward the construction of two tests (A and B), each containing one of the two compared SAQs.

As shown in Figures 2 and 3, each of the SAQ diagrams incorporates a group of organic compounds (concepts) that are, or can be, interconnected via chemical transformations (reactions) forming a basic closed loop. Some of these compounds are also chemically related with other compounds not belonging to the basic loop. Furthermore, in the diagram of the second SAQ, which includes a relatively large number of compounds, a more complex net of connections is created leading to the formation of conceptual subloops inside the external basic loop.

Second, we had to select the amount of what would be provided to and what would be required from the examinees. Once again, we selected different amounts of provided features and requirements for each of the two SAQs, with the second SAQ being more demanding for the examinees. Moreover, we decided that the second SAQ would contain one additional subquestion (B1a in Figure 3). In this way, the second SAQ was constructed to be "less-directed" compared with the first one. The more demanding and "less-directed" SAQ was part of the second administered test because by this time students should be more familiar with SAQs.

Our next concern was to establish a clear scoring method for SAQ items. In a SAQ's systemic diagram, the various components of the diagram are all interrelated, directly or indirectly, constructing a meaningful whole, namely, an interconnected conceptual system, with a nonhierarchical structure. We consider

B1. In the following diagram:

a. Define the direction of the undirected linking lines (4 lines).

b. Fill in the 5 blank squares with the chemical formulas of the proper compounds.

c. Fill in the 5 blanks on the arrows with the reagents/conditions of the reactions.

d. Fill in two more chemical reactions between compounds in the diagram. (Draw the two arrows corresponding to those reactions and label the arrows with the names and types of the reactions.)

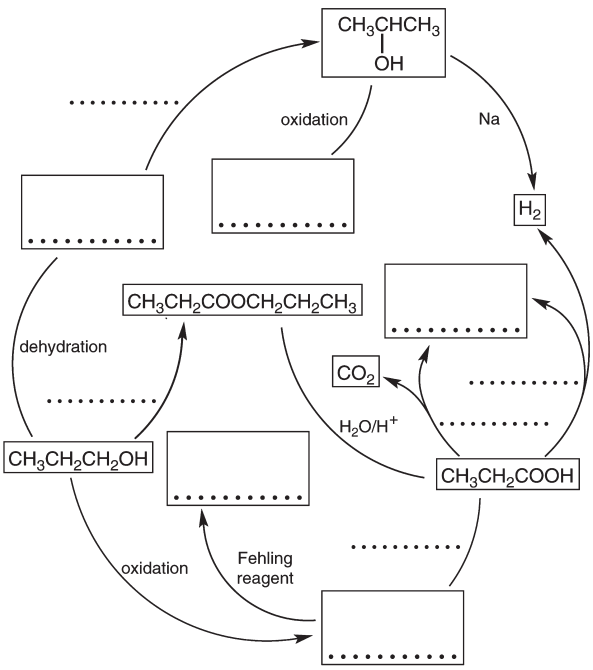

Figure 3. SAQ of the second (B) test.

A2. Fill in the blanks in the following propositions:

a. By the addition of water in acetylene in the presence of $H_2SO_4/HgSO_4$, the final stable product is ________________ .

b. ________________ is produced by the trimerization of acetylene.

A3. Fill in the blanks in the following chemical reactions:

a. $CH_3C{\equiv}CH + HCl \longrightarrow \ldots\ldots\ldots \xrightarrow{+\ HCl} \ldots\ldots\ldots$

b. $CH_3C{\equiv}CH + Na \longrightarrow \ldots\ldots\ldots$ (the organic product)

A4. Choose the right answer for each of the propositions, a and b:

a. By the two-stage addition of hydrogen to acetylene, which product is formed?
i. ethane ii. ethene
iii. propane iv. propene

b. For two compounds, *X* and *Y*, the following information is known:
- They both decolorize a solution of bromine in tetrachloromethane.
- Only compound *X* reacts with sodium.
- By addition of H_2 in compound *X* the compound *Y* is formed.

Accordingly, which of the following is correct?

i. Compound *X* is propene and Compound *Y* is propine.
ii. Compound *X* is propine and Compound *Y* is propene.
iii. Compound *X* is propane and Compound *Y* is propene.
iv. Compound *X* is ethene and Compound *Y* is acetylene.

Figure 4. Conventional questions of the first (A) test.

that each of these components equally contributes to the creation of this conceptual whole. Therefore, we suggested a one-point score for each valid component filled-in, be it a concept, directed linking line (arrow), linking phrase, or direction of a provided undirected linking line. This scoring method focuses primarily on the accuracy of the content.

Each of the two tests also contained some conventional objective questions (Figures 4 and 5) that would allow us to collect evidence, through the conduction of exploratory factor analysis, for the "meaningful—rote" character of SAQs under study. The factor analysis is a statistic method providing evidence for construct validity *(33)*.

The conventional questions were constructed to assess simple recall of knowledge, except for one question (A4b in Figure 4), which was assumed to demand higher cognitive skills. We incorporated this question in order to see whether, and in which extent, a more demanding conventional question would be related with SAQ items. The types of conventional questions in both tests were "fill-in-the blanks" and multiple choice and they were scored one point for each valid answer (data point). Although conventional questions were purposefully constructed to be closer to the "rote" edge of the learning continuum, we were also interested in observing the influence of some other variables on their "meaningful—rote" character. One of these variables was the "sequential" form of organic reactions. What happens when we have two sequential reactions in which the product of the first is reactant for the second reaction? Our experience in teaching high school organic chemistry as well as a report in the literature *(16)* suggest that students are often incapable of relating individual reactions in sequential reaction schemes. Therefore, we included in the tests some conventional questions having a "sequential reaction" format (A3a in Figure 4, B3a and B3d in Figure 5). Furthermore, some of the employed conventional questions included organic compounds that were not exactly the same as those presented in the textbook. This variable is related with transfer of knowledge and was expected to enhance, more or less, the questions' "meaningful" character.

Taking under consideration the 11th grade Greek chemistry curriculum, we constructed the first test (Figures 2 and 4) to assess students' understanding of hydrocarbon chemistry, while the second test (Figures 3 and 5) assessed their understanding of the chemistry of alcohols and carboxylic acids. Care was taken so that appropriate language and representations were used. The tests were discussed for *content validity* with two experienced secondary education teachers and this was followed by consultation with a chemical educator and a professor from the Department of Chemistry, University of Athens. Both tests were subsumed under the formal high school two-semester summative assessment. The required time for answering each test was within one Greek high school teaching period (40 min).

Participants

The participants in this study were 72 11th grade high school students from three intact classes in a public suburban high school in the Athens area. Because not all students were present on the days of the data collection, 65 (35 males, 30 females) out of 72 students completed the first test, while 49 students (32 males, 17 females) completed the second one.

Procedure

The study was conducted over a 5-month period (December to April) and in two stages (Table 1): In the first stage, the chemistry of hydrocarbons was taught. This content unit was taught over 10 class periods (40 min each) by one chemistry teacher using the traditional approach (lectures with presentations and discussions). During these lessons, students were provided with worksheets that, in the first place, included various types of conventional objective questions as well as some linear questions having similar format with SAQs but different forms of diagrams. Linear questions, in contrast with SAQs, included diagrams with concepts in a linear arrangement. The purpose of using these linear questions was so that the students could become familiar with the symbolic representation of organic reactions used in SAQs. As instruction on the topic was underway, students were provided with worksheets that included some additional authentic SAQs. After completing the teaching—learning procedure, the teacher sought feedback from the students: that information assured the teacher that the students had really understood the SAQ scheme, obtaining a satisfactory degree of familiarity with this type of questions. This was also verified after a thorough overview of the worksheets completed by the students was carried out. At the end of this stage the first test was administered.

In the second stage, the other two organic chemistry topics (chemistry of alcohols and carboxylic acids) were taught. This content unit was taught over 12 class periods by the same chemistry teacher, using two series of systemic diagrams (five diagrams in total) corresponding to the two topics under study. On the basis of the preliminary diagram from each series, the students, guided and supported by their teacher, used a step-by-step approach to construct the final full diagram for each topic. In addition, students were provided with worksheets including various SAQ formats and some conventional questions. The last step of this stage was the administration of the second test.

According to the Greek high school chemistry curriculum, 12 class periods should be dedicated for the instruction of the topics under study. However, in this study, additional instructional time was required for the students to understand the systemic philosophy of SAQs. Nevertheless, we expect the requested instructional time to be progressively reduced as students become more familiar with the SAQ scheme. After collection, the data were analyzed using statistics procedures and statistical analysis software (SPSS 11.0). The results are presented and discussed below in detail.

B2. Choose the right answer for each of the following propositions:

a. It cannot change the color of an acid solution of $K_2Cr_2O_7$.

i. ethanol ii. propanone
iii. 2-propanol iv. propanal

b. It shows acid character.

i. ethanal ii. 1-propanol
iii. butanone iv. propanone

c. It can be detected with addition of a carbonate salt.

i. 2-butanol ii. butanal
iii. butanoic acid iv. ethanol

B3. Fill in the blanks:

a. By oxidation of ethanol is firstly formed _________ and then _________ .

b. By the reaction of ethanoic acid with methanol in an acid environment, _________ is formed.

c. $CH_3CN + H_2O \xrightarrow{H^+ \text{ or } OH^-} \ldots\ldots\ldots + \ldots\ldots\ldots$

d. $CH_3CH{=}CH_2 + H_2O \xrightarrow{H^+} \ldots\ldots\ldots$

$\xrightarrow{+[O]} \ldots\ldots\ldots + H_2O$

Figure 5. Conventional questions of the second (B) test.

Results and Discussion

For the statistics procedures, we summed the scores from the data points belonging to the same subquestion into a total score for each subquestion. We did that considering that the data points from the same subquestion are interrelated and consequently they should be processed statistically as a whole. As a result, the statistic procedures were conducted on 9 subquestions' scores for the first test and 11 subquestions' scores for the second. From now on in this paper, these subquestions will be called "items".

The evidence of *items' validity* was calculated with "item—total score" correlations using the Pearson's correlation coefficient (r) (Table 2). For the first test, two items out of nine were

Table 1. Procedure for the Study

Stage	Phase	Duration	Teaching Method
1. Chemistry of Hydrocarbons	Teaching-learning procedure	10×40 min.	Traditional (lectures, demonstrations, discussions)
	Administration of first test	40 min.	
2. Chemistry of Alcohols and Carboxylic Acids	Teaching-learning procedure	12×40 min.	Implementation of SATL techniques
	Administration of second test	40 min.	

described as appropriate ($0.30 < r < 0.39$), while the remaining seven as quite appropriate ($r > 0.40$). For the second test, two items out of 11 were described as marginally appropriate ($r = 0.20$), one item as appropriate ($0.30 < r < 0.39$) and the remaining eight as quite appropriate ($r > 0.40$). These results indicated that all items contribute to the validity of the tests.

The *reliability* for each test was calculated by two different techniques: (i) split-half reliability and (ii) reliability of "internal consistency". The results for both tests, listed in Table 3, show that the scales have an acceptable reliability. The relatively low split-half reliability for part two of the first test could be explained by the small number of included items.

The hypothesis of the two dimensions ("meaningful" versus "rote") was tested using exploratory factor analysis. For the first test, the correlation matrix for the 9 items was computed. The Bartlett's test of sphericity gave a value of 147.034 with a significance level of $p < 0.001$, indicating the suitability of the factor model for the data under consideration. The Kaiser–Meyer–Olkin measure of sampling adequacy was also used. The obtained value (0.812) indicated the appropriateness of the factor model as well. Principal component analysis resulted in two common factors with eigen values >1, which all together explain 53% of the variance. The result of the two common factors was in accordance with our theoretical considerations. The two factors were subjected to a varimax rotation. A minimum factor-loading criterion of 0.40 *(34)* was adopted for the final interpretation of the results. Table 4 reports the rotated factor loadings.

Items A1a, A1b, A2a, A2b, A3a, A3b, and A4a, were loaded on the first factor, while items A1c and A4b on the second. Taking into consideration the items' content and requirements, the more reasonable explanation is that the first principal factor is the "rote" factor, while the second is the "meaningful" one. Items A1c and A4b seem to demand higher-order cognitive skills from the examinees compared to the remaining items. In item A1c, which is a SAQ item, students are asked to provide two additional chemical relationships (reactions) in the given systemic diagram, that is, four new components of the systemic diagram are required. In item A4b, students have to take into consideration three chemical properties of two given unknown compounds in order to identify them. These two items, as was expected, were strongly loaded on the "meaningful" factor. On the other hand, the remaining two SAQ items (A1a and A1b) were both strongly loaded on the "rote" factor, with item A1a

Table 2. Item–Total Score Correlations

First Test		Second Test	
Item	*r* Values	Item	*r* Values
A1a[a]	0.64	B1a[a]	0.50
A1b[a]	0.53	B1b[a]	0.81
A1c[a]	0.46	B1c[a]	0.75
A2a	0.48	B1d[a]	0.42
A2b	0.43	B2a	0.38
A3a	0.61	B2b	0.20
A3b	0.54	B2c	0.20
A4a	0.39	B3a	0.76
A4b	0.35	B3b	0.68
		B3c	0.40
		B3d	0.64

[a] SAQ items.

Table 4. Rotated Component Matrices of Tests' Items

First Test (A)			Second Test (B)		
	Component			Component	
Item	1	2	Item	1	2
A2a	**0.75**	−0.04	B1b[a]	**0.86**	0.24
A1b[a]	**0.70**	0.10	B1c[a]	**0.85**	0.11
A1a[a]	**0.70**	0.33	B3a	**0.73**	**0.44**
A3b	**0.69**	0.16	B3b	**0.67**	**0.44**
A3a	**0.64**	0.39	B1a[a]	**0.65**	−0.03
A2b	**0.61**	0.04	B3d	**0.60**	**0.53**
A4a	**0.51**	0.18	B2a	**0.57**	−0.15
A4b	0.00	**0.87**	B1d[a]	**0.55**	−0.03
A1c[a]	0.27	**0.75**	B3c	**0.40**	0.39
			B2b	−0.08	**0.77**
			B2c	−0.02	**0.71**

[a] SAQ items. Note: Loadings equal or greater than 0.40 are indicated in bold.

Table 3. Reliability Calculations for the SAQ Tests

Measures and *r* Values, First Test		Measures and *r* Values, Second Test	
Split-Half Methods To Assess Reliability			
Correlation between Forms	0.69	Correlation between Forms	0.66
Guttman Split-Half	0.72	Guttman Split-Half	0.69
Alpha for Part 1 (5 Items in Part 1)	0.64	Alpha for Part 1 (6 Items in Part 1)	0.69
Equal-Length Spearman–Brown	0.81	Equal-Length Spearman–Brown	0.79
Unequal-Length Spearman–Brown	0.81	Unequal-Length Spearman–Brown	0.79
Alpha for Part 2 (4 Items in Part 2)	0.57	Alpha for Part 2 (5 Items in Part 2)	0.80
Reliability of "Internal Consistency"			
Cronbach's Alpha	0.76	Cronbach's Alpha	0.83

giving simultaneously a noticeable loading on the "meaningful" factor (0.33). Furthermore, the conventional items, except for A4b, were strongly (A2a, A2b, A3a, A3b) to moderately (A4a) loaded on "rote" factor, as expected, while item A3a also gave a quite noticeable loading on "meaningful" factor (0.39).

In the first SAQ, the fact that two of the three items (A1a and A1b) were strongly loaded on the "rote" factor indicates that the characteristics of this SAQ are not suitable for assessing meaningful learning. Perhaps the corresponding systemic diagram did not have the proper complexity, or what was required from the examinees was not sufficiently compared to the amount and significance of what was provided to them. Only the third SAQ item (A1c), which required students to provide two additional chemical relationships between compounds in the systemic diagram, was loaded on the "meaningful" factor. Thus, it seems that this specific item is suitable for capturing meaningful learning, independently from the SAQ's characteristics. This result is in line with the previously reported and above mentioned *(30)*, given that item A1c is individually more "less-directed" compared to the remaining SAQ items. On the other hand, item A1a, although strongly loaded on the "rote" factor, just like item A1b was, seemed to be related up to a point with the "meaningful" factor as well.

This fragmentation of SAQ items into "rote" and "meaningful" can be explained by conceptual change theory. According to the initial model of conceptual change *(35)*, there are two phases of conceptual change in learning: (i) assimilation, regarding the use of existing concepts to deal with new phenomena, and (ii) accommodation, regarding more radical changes of the conceptual structures, namely, students must replace or reorganize their central concepts. Other contributions or alternative models have been suggested afterward (e.g., 36−38)). All the proposed models recognize two possible kinds of changes to cognitive structures during the learning procedure: Changes that involve the simple addition of knowledge (e.g., accretion), or, alternatively, changes that involve revisions of the existing conceptual structures rather than simple addition. The latter kind of learning is most commonly described as conceptual change and is divided into strong revision (e.g., accommodation) and weak or lesser revision (e.g., assimilation) by most theorists. Creating links is an important feature of conceptual change theory; otherwise there is no difference between conceptual change and simple rote learning *(39)*. In this reasoning, items A1a and A1b in our study, asking the students to complete concepts or linking phrases in structured relationships, seem to be more related with accretion, while item A1c, requiring from the students to provide additional links between concepts, should be closer to assimilation.

For the conventional items in the first test, the results were according to our "simple recall of knowledge" assumptions, except for item A3a, which showed a relatively increased "meaningful" character. This could be due to at least two reasons: the first is the presence of two sequential reactions in item A3a, which probably makes this item a little more "meaningful". The second reason is that the addition of HCl to propine according to Markovnikov's rule is not equivalent with the addition to acetylene, which is the reaction referred in the textbook. Using this reasoning, a degree of knowledge transfer is demanded for answering item A3a, which could justify the observed loading on the "meaningful" factor.

For the second test, the Bartlett's test of sphericity for the 11 items gave a value of 223.872 with a significance level of $p < 0.001$. For the Kaiser−Meyer−Olkin measure of sampling adequacy the obtained value was 0.788. Principal component analysis resulted in three common factors with eigen values >1, which all together explain 65% of the variance. On the basis of scree-plot and theoretical considerations, we used two of the three factors, accounting for 55% of the variance, which were subjected to a varimax rotation. The rotated factor loadings are presented in Table 4. Items B1a, B1b, B1c, B1d, and B2a, were loaded on the first principal factor, items B2b and B2c on the second, while items B3a, B3b, and B3d, on both factors, though stronger on the first. Finally, item B3c was almost equally loaded on both factors, right on the cutoff criterion used (0.40). Once again taking under consideration the items' content and requirements, it seems that the first factor is the "meaningful", while the second is the "rote".

All the items of the second SAQ were strongly (B1a, B1b, B1c) to moderately (B1d) loaded on the "meaningful" factor. This result indicates that the characteristics of this SAQ are suitable for assessing meaningful learning and definitely more suitable than those of the SAQ in the first test. A logic conclusion is that the more "less-directed" SAQ, including a more complex systemic diagram and with higher demands for the examinees (SAQ in the second test), was found to be more appropriate for capturing students' meaningful understanding of organic reactions. This conclusion is consistent with reported results *(30)*, which revealed that between two compared concept mapping techniques, the more "less-directed" and demanding was the one that better reflected differences among students' knowledge structures.

From conventional items, B2a, although it was assumed to assess simple recall of knowledge, was found to be more "meaningful" having a moderate loading on "meaningful" factor (0.57). Perhaps this is due to the way it was enounced, which is a little different from the way this theme is presented in the textbook. The question "which compound cannot change the color of an acid solution of $K_2Cr_2O_7$?" probably does not demand just recall of knowledge. According to the textbook, an alternative construction "which compound cannot be oxidized by an acid solution of $K_2Cr_2O_7$?" could be more "rote". On the other hand, items B2b and B2c were strongly loaded on "rote" factor as was expected.

Items B3a, B3b, and B3d, were loaded on both factors (strongly on "meaningful" factor while moderately on "rote" factor). An explanation for the increased "meaningful" loadings of items B3a and B3d might be that, like item A3a, they include two sequential reactions. Moreover, for item B3d, the addition of water to propene according to Markovnikov's rule is not equivalent with the addition to ethene mentioned in the textbook. Furthermore, for item B3a, the textbook states, "Primary alcohols are oxidized to aldehydes, and aldehydes are oxidized to carboxylic acids." As a result, some students might have not clearly understood that acetic acid can be directly provided by ethanol oxidation via the in situ formation of ethanal, as item B3a implies. The loading of item B3b on "meaningful" factor could be explained by the fact that the item is concerned with esterification of methanol and not of ethanol, which is the compound mentioned in the textbook, although these two reactions are equivalent. Additionally, item B3c was marginally and almost equally loaded on both factors ("meaningful" 0.40

and "rote" 0.39). In any case, the fact that some items were loaded on both factors ("meaningful" and "rote") is something that is theoretically acceptable. Meaningful learning is not an absolute, but a trait of learning measured by degree. Rote and meaningful learning are not dichotomous and mutually exclusive. They are parts of the same continuum and sometimes they can even occur simultaneously *(40)*. Finally, our "meaningful—rote" interpretation is also supported by the fact that items' loadings in factor analysis were not determined only by items' difficulty. Namely, some items that were loaded on the "meaningful" factor had similar difficulty with other items in the same test, which were loaded on the "rote" factor.

Conclusion

This investigation indicates that SAQs under study have acceptable psychometric properties and are suitable to be used as assessment tools in high school. Exploratory factor analysis revealed that the characteristics of SAQs seem to play a significant role in their effectiveness for assessing meaningful learning. Between the two compared "fill-in-the blanks" SAQs, the more "less-directed" and demanding, which incorporates a more complex systemic diagram, was found to be more suitable for this purpose. Concerning the conventional questions constructed to assess simple recall of knowledge, their "sequential reactions" format as well as the incorporation of "not included in the textbook" compounds seem to enhance their "meaningful" character. We believe that these results indicate a pathway for a more thorough classroom assessment of organic reactions. As it has been suggested *(31)*, the term "assessment" reflects the belief that reaching a judgment about an individual's achievement in a domain requires an integration of several pieces of information. Assessment based on concept maps is only one of those pieces. Furthermore, properly written subjective or multiple-choice tests are considered indispensable in assessing what students learn and help teachers to understand students' thinking *(41)*. The results that emerged from this study are consisted with these considerations. It seems that a combination between SAQs with specific characteristics and properly designed conventional objective questions could capture more efficiently high school students' meaningful understanding of organic reactions. For example, students could be asked to use a reaction sequence in the SAQ scheme for answering an additional explanation question. Thus, the instructor would have further evidence that they actually understood the SAQ scheme.

In the future, it would also be interesting to explore other, more "less-directed", SAQ types, for example, the case of student-generated SAQs. At this point, an observation coming from an anonymous reviewer regarding a student's performance in college-level organic chemistry should be mentioned. This student demonstrated the most improvement from his class exam average score to his score on the final exam. When the instructor mentioned this fact to him, he showed a "map" of organic reactions he had constructed in which he included as many reactions as he could in a logical scheme *(42)*. This observation further indicates that student-generated SAQs are worth investigating. In such a case, care must be taken to ensure that the characteristics and requirements of SAQ formats are suitable for assessment in secondary education. Additionally, the scoring method for these "open-ended" SAQ types should be carefully addressed.

In classroom settings, the focus should be on helping students understand rather than memorize chemistry content. Chemistry teachers should choose instructional approaches that encourage students to learn meaningfully, developing a more coherent conceptual understanding. The chemistry curricula content would be more useful if regarded as a vehicle to enhance and improve students' thinking rather than a quantity to be memorized. We believe that the implementation of SATL techniques could be a promising alternative, as it offers a systemic perspective of the subject matter. Regarding assessment, more emphasis should be dedicated to assessing students' deep understanding of chemistry concepts. It seems that the SAQ scheme, by encouraging and promoting the development of systems thinking, can facilitate students to integrate concepts in meaningful wholes and thus to accomplish a deeper conceptual understanding of organic reactions. Furthermore, by using the SAQ assessment, a matching is attained between teaching—learning procedures and assessment within SATL model. Thus, students' learning is assessed in the major degree in the context of teaching, making assessment more "authentic". The latter is one of the principles that should guide teachers in implementing constructivist ideas in classroom settings *(43)*.

One limitation of this study is that the SAQs under investigation are concerned only with the symbolic level of chemistry, as it is expressed by chemical formulas and equations. A challenge for future research would be the evaluation of various SAQ forms for assessing students' meaningful understanding in other chemistry levels (submicroscale and macroscale) as well as in other chemistry topics, besides organic reactions. In this direction, further evidence for the validity and reliability of SAQs would be very useful, allowing a more extensive use of these tools for classroom assessment and for an effective evaluation of the SATL model as well.

Acknowledgment

Theodoros Vachliotis, graduate student, gratefully acknowledges the State Scholarships Foundation of Athens, Greece, for financial support.

Literature Cited

1. Fahmy, A. F. M.; Lagowski, J. J. *Pure Appl. Chem.* **1999**, *71*, 859—863. http://media.iupac.org/publications/pac/1999/pdf/7105x0859.pdf (accessed Dec 2010).
2. Fahmy, A. F. M.; Lagowski, J. J. *Chem. Educ. Int.* **2002**, *3*, 1. http://old.iupac.org/publications/cei/vol3/0301x0an1.html (accessed Dec 2010).
3. Fahmy, A. F. M.; Lagowski, J. J. *J. Chem. Educ.* **2003**, *80*, 1078–1083.
4. Lagowski, J. J. *J. Chem. Educ.* **2005**, *82*, 211.
5. Lagowski, J. J. *SATL, Learning Theory, and the Physiology of Learning*; Paper presented at the 20th International Conference on Chemical Education, August 3—8, 2008, Mauritius. http://www.satlcentral.com/e-conferences.htm (accessed Dec 2010).
6. Fahmy, A. F. M.; Lagowski, J. J. Using SATL Techniques To Assess Student Achievement. In Proceedings, 18th International Conference on Chemical Education, August 3—8, 2004, Istanbul, Turkey, S14.1.
7. Fahmy, A. F. M.; Lagowski, J. J. *Chem. Educ. Int.* **2007/08**, *8*, 1. http://old.iupac.org/publications/cei/vol8/0801xFahmy.pdf (accessed Dec 2010).

8. Schwaninger, M. *Syst. Res.* **2006**, *23*, 583–594.
9. Dyehouse, M.; Bennett, D.; Harbor, J.; Childress, A.; Dark, M. *Eval. Program Plann.* **2009**, *32*, 187–196.
10. Midgley, G. Systems Thinking for Evaluation. In *Systems Concepts in Evaluation: An Expert Anthology*, Williams, B., Imam, I., Eds.; EdgePress of Inverness: Point Reyes, CA, 2007; pp 11−34.
11. Cabrera, D.; Colosi, L.; Lobdell, C. *Eval. Program Plann.* **2008**, *31*, 299–310.
12. Cabrera, D. A. *Systems Thinking*. Doctoral dissertation, Cornell University; Dissertation Abstracts International−A, 67, 303, 2006.
13. Ausubel, D. P. *The Psychology of Meaningful Verbal Learning*; Grune and Stratton: New York, 1963.
14. Novak, J. D. *Sci. Educ.* **2002**, *86*, 548–571.
15. Hassan, A. K.; Hill, R. A.; Reid, N. *U. Chem. Ed.* **2004**, *8*, 40–51.
16. Šket, B.; Glažar, S. A. *Acta Chim. Slov.* **2005**, *52*, 471−477. http://acta.chem-soc.si/52/52-4-471.pdf (accessed Dec 2010).
17. Schmidt, H. I. *Adv. Health Sci. Educ.* **2006**, *11*, 69–76.
18. Michael, J. *Adv. Physiol. Educ.* **2001**, *25*, 145–158.
19. Mayer, R. E. *Theor. Pract.* **2002**, *41*, 226–232.
20. Novak, J. D.; Mintzes, J. J.; Wandersee, J. H. Epilogue: On Ways of Assessing Science Understanding. In *Assessing Science Understanding: A Human Constructivist View*; Mintzes, J. J., Wandersee, J. H., Novak, J. D., Eds.; Academic Press: San Diego, CA, 2000; pp 355−374.
21. Stoddart, T.; Abrams, R.; Gasper, E.; Canaday, D. *Int. J. Sci. Educ.* **2000**, *22*, 1221–1246.
22. Novak, J. D.; Gowin, D. B. *Learning How To Learn*; University Press of Cambridge: Cambridge, 1984.
23. Ruiz-Primo, M. A.; Shavelson, R. J. *J. Res. Sci. Teach.* **1996**, *33*, 569–600.
24. Shaka, F. L.; Bitner, B. L. *Construction and Validation of a Rubric for Scoring Concept Maps*. Paper presented at the Association for the Education of Teachers of Science, 1996, Seattle, WA.
25. Kinchin, I. M.; Hay, D. B.; Adams, A. *Educ. Res.* **2000**, *42*, 43–57.
26. Shavelson, R. J.; Ruiz-Primo, M. A. On the Psychometrics of Assessing Science Understanding. In *Assessing Science Understanding: A Human Constructivist View*; Mintzes, J. J., Wandersee, J. H., Novak, J. D., Eds.; Academic Press: San Diego, CA, 2000; pp 304−341.
27. McClure, J. R.; Sonak, B.; Suen, H. K. *J. Res. Sci. Teach.* **1999**, *36*, 475–492.
28. West, D.; Pomeroy, J. R.; Park, J.; Gerstenberger, E.; Sandoval, J. *J. Amer. Med. Assoc.* **2000**, *284*, 1105–1110.
29. Novak, J. D.; Gowin, D. B.; Johansen, G. *Sci. Educ.* **1983**, *67*, 625–645.
30. Ruiz-Primo, M. A.; Schultz, S. E.; Li, M.; Shavelson, R. J. *J. Res. Sci. Teach.* **2001**, *38*, 260–278.
31. Ruiz-Primo, M. A. Examining Concept Maps as an Assessment Tool. In *Proceedings of the First International Conference on Concept Mapping*, September 14−17, **2004**; Cañas, A. J., Novak, J. D., González, F. M., Eds.; Pamplona, Spain, **2004**.
32. Miller, G. *Psychol. Rev.* **1956**, *63*, 81–97.
33. Anastasi, A.; Urbina, S. *Psychological Testing*, 7th ed.; Prentice-Hall, Inc.: Upper Saddle River, NJ, 1997.
34. Stevens, J. P. *Applied Multivariate Statistics for the Social Sciences*, 4th ed.; Erlbaum: Mahwah, NJ, 2002.
35. Posner, G. J.; Strike, K. A.; Hewson, P. W.; Gertzog, W. A. *Sci. Educ.* **1982**, *66*, 211–227.
36. Vosniadou, S. *Learn. Instr.* **1994**, *4*, 45–69.
37. Chi, M. T. H.; Slotta, J. D.; deLeeuw, N. *Learn. Instr.* **1994**, *4*, 27–43.
38. diSessa, A. A. *Cognition Instr.* **1993**, *10*, 105–225.
39. Tyson, L. M.; Venville, G. J.; Harrison, A. G.; Treagust, D. F. *Sci. Educ.* **1997**, *81*, 387–404.
40. Stefani, C.; Tsaparlis, G. *J. Res. Sci. Teach.* **2009**, *46*, 520–536.
41. BouJaoude, S.; Barakat, H. *Electronic J. Sci. Educ.* **2003**, *7* (No. 3). http://wolfweb.unr.edu/homepage/crowther/ejse/boujaoude.pdf (accessed March **2010**).
42. Observation coming from an anonymous reviewer.
43. Brooks, G.; Brooks, M. *In Search of Understanding: The Case for Constructivist Classrooms*; Association for Supervision and Curriculum Development: Alexandria, VA, 1999.