

Student recognition and construction of quality chemistry essay responses

Diane M. Bunce* and Jessica R. VandenPlas

Department of Chemistry, The Catholic University of America, Washington, DC 20064
e-mail: bunce@cua.edu

Received 16 January 2006, accepted 24 April 2006

Abstract: Students in chemistry traditionally experience more difficulty responding to essay questions than to calculating a numerical answer for the same concept. The purpose of this study is to investigate students' understanding of what constitutes a complete and cogent essay answer in chemistry. Preliminary data from thirty-nine non-science majors support the hypothesis that students do not reliably recognize or construct adequate responses to chemistry essay questions. In addition, students' intention to construct a complete and cogent argument is compared to their actual responses. Inability to construct complete and cogent arguments may result in lower achievement scores on essay questions. Since essay questions are typically used to test both achievement and the effectiveness of innovative teaching practices, this situation may mask significant research results. These data suggest the need for a more extensive investigation of student construction of quality essay answers in chemistry as well as in other sciences. [*Chem. Educ. Res. Pract.*, 2006, 7 (3), 160-169]

Keywords: assessment, open-ended questions, cognitive load, essay questions, research methodology

Introduction

In many large university general chemistry courses, multiple choice and numerical application problems have traditionally been the assessment method of choice due to ease of administration, but such questions do not adequately measure student understanding (Moore, 1997). Student understanding can be explained in terms of Ausubel's continuum between meaningful learning and rote learning (Novak and Gowin, 1984). Rote learning lends itself more easily to being tested by multiple choice questions, while meaningful learning is better assessed through essay-type questions. Research also shows that students do not perform as well on essay or short answer test items as they do on multiple choice test items (Danili and Reid, 2005). In spite of this fact, essay questions are often used as a means of evaluating the effectiveness of innovative teaching methods which stress meaningful learning (Bodner, 1991; Oliver-Hoyo et al., 2004). Since both assessment of student understanding and evaluation of the effectiveness of teaching innovations depend on the quality of student responses to essay questions, it becomes important to ascertain why students are not performing as well on these questions. The problem may be either that students don't know the science to answer the question or don't realize the logic necessary to construct an adequate answer. The purpose of this study is to investigate this problem in terms of whether students can recognize and/or construct complete and cogent responses to essay questions in chemistry.

Inherent in the assessment of students' understanding of chemistry is the cognitive demand of both the subject of chemistry and the way chemistry questions are asked. Cognitive Load Theory (Sweller, 1994) identifies two factors of information processing that should be taken into consideration in student assessment: 1) the nature of the subject being assessed (intrinsic variable) and 2) how the assessment is designed (extraneous variable).

In chemistry, the magnitude of the intrinsic variable is high due to the fact that there is much interaction among the elements of each concept. For instance, it is difficult to study bonding if you don't understand molecular polarity. Molecular polarity, as predicted by VSEPR theory, requires in turn knowledge of Lewis Dot diagrams. Using Lewis Dot diagrams is based upon an understanding of electron configuration, which in turn is based upon an understanding of atomic structure. So, a seemingly 'direct' question on bonding requires knowledge of at least five other chemistry concepts and empirical measures. This inter-relatedness of concepts and empirical measures increases the intrinsic variable of answering chemistry questions regardless of the assessment format.

The second factor of Cognitive Load Theory is the extraneous variable, which deals with how the assessment is designed. For instance, there are several ways to ask a density problem on a test including determination of a numerical answer and/or explanation based upon the chemistry concept. In the simplest form, a student can be asked to calculate the density, given the mass and volume. Here the student essentially 'plugs and chugs' through the problem. Understanding of the concept is not as important in this situation as the application of a numerical algorithm. The extraneous load can be increased by asking the student to explain why something happens, such as why lead is more dense than iron. Here, the student must present both the underlying concept of density and explain it in a logical argument. As the extraneous variable increases, fewer students can determine the correct answer.

One advantage of using essay questions in chemistry is that the process of answering them helps students organize their information in long term memory (Wandensee et al., 1994). The ability to organize information in long term memory is critical to the transition from novice to expert understanding. Experts effectively organize their knowledge in long term memory, and this organization facilitates their ability to learn new information. Novices, on the other hand, do not organize their knowledge effectively in long term memory. As a result, their knowledge is often fragmented and thus does not help them learn additional information. Novices can be supported in their reorganization of knowledge through practice in planning responses to essay questions.

In assessment using essay questions the emphasis is not only on *what* happens but also on *why* it happens. This change should help shift the focus from students memorizing information to their placing emphasis on truly understanding the concepts and the way in which these concepts are related. Such a shift should increase the organization of information in long term memory, thus aiding in the transition from novice to expert. An added benefit of using essay questions is that they can be used by the teacher as a formative evaluation, which helps to reveal where students experience a lack of understanding or misconceptions about a chemistry concept.

Once the decision is made to include essay questions in chemistry, some thought must be given to the structure of such questions. The need to keep directions simple and avoid overloading the short term memory capacity of the student is essential (Cavallo et al., 2003). In addition, the amount of information that the student must process simultaneously should be kept to a minimum. The inclusion of key terms that serve as 'anchors' for the schemas that students used to store the original information will help ensure that students address the question asked in the way the teacher has planned.

Even if essay questions are constructed in keeping with Cavallo's theory, students still need specific instruction and practice in addressing them. Such instruction includes allowing

students access to the objective and explicit rubric used to grade their answers, as well as timely feedback so that students can learn from their mistakes (Kovac and Sherwood, 1999). Experience with teaching essay writing in chemistry (Russell, 2004) has shown that students need multiple exposures to both critiquing and writing chemistry essays in order to develop the skill of producing persuasive and logical arguments.

In this study it was hypothesized, based upon Russell's (2004) research, that if students had practice evaluating sample answers to essay chemistry questions in terms of completeness and cogency, they would be better able to construct quality answers to the same questions on a separate occasion. It was further hypothesized that the opportunity to plan an answer to such questions would improve student responses by providing them with an opportunity to organize their knowledge in a more expert fashion. Additional variables of logical reasoning ability and year in school were also investigated for their effect on the construction of quality answers to chemistry essay questions.

Research questions

Complete essay responses are those that clearly and fully address the questions asked, while cogent responses use a logical argument as a means of answering the question. Quality answers to questions should be both complete and cogent, but student answers often are not. Less ideal answers can be complete but not cogent, cogent but not complete, or neither complete nor cogent.

The following research questions were investigated in this study:

- 1) Can students *recognize* a complete and cogent response to a chemistry essay question?
- 2) Can students *construct* a complete and cogent response to a chemistry essay question?
- 3) Does the practicing of evaluating sample essay responses for completeness and cogency increase student ability to construct quality answers to the same questions at a later time?
- 4) Does a relationship exist between students' planned and actual responses to chemistry essay questions?

Methodology

Sample

The sample for this study was thirty-nine undergraduate students (first through fourth year) enrolled in a chemistry course for non-science majors at a small private university in the Mid-Atlantic region of the USA. The experimental design of this study was reviewed by the University to ensure compliance with Protection of Human Subjects Protocol.

Essay questions

The questions in this study were essay questions used on exams from a previous year. They included questions on the topics of infrared absorption, interpretation of IR spectra, gasoline octane ratings and applications of entropy. Sample questions are shown in Figure 1.

Figure 1. Sample essay questions.

Exercise 4-1 (Gasoline). You and your younger brother fill up the family car with fuel on vacation. Your brother asks if there is really any difference besides price between regular and premium gasoline. Based on what you learned in this course, what would you tell him? Be sure to include as much chemistry as you can (just to impress him).

Exercise 4-3 (Entropy). Entropy is involved in many of the things you do in real life. For instance, in cooking when you melt a stick of butter, entropy is involved. Explain what happens in terms of entropy when solid butter is melted. Support your answer.

Instruments

Group Assessment of Logical Thinking (GALT) test

It was hypothesised that students' logical thinking ability would influence their ability to construct cogent essay responses. The Group Assessment of Logical Thinking (GALT) test (Roadrangka et al., 1982) was therefore used to measure logical reasoning ability. This test consists of 12 questions, each of which involves a correct answer and a reason for that answer. GALT scores range from 0 to 12. An online version of this test was used in this study.

A frequency distribution of scores was used to determine natural cut-off points for high (9-12), medium (7-8) and low (0-6) GALT levels in this sample. Due to the small number of students in the medium GALT level, this level was excluded from analysis. The number of students in the remaining GALT levels is given in Table 1.

Online survey

It was further hypothesised that students' ability to write complete and cogent essay responses would improve with their advancement in the University as measured by their class level. Demographic data, including undergraduate class level, were collected via an online survey. Table 1 summarizes the number of students in upper (third and fourth year undergraduate) and lower (first and second year undergraduate) class levels.

Table 1. Number of students by GALT and class level.

Characteristics of Subjects	Levels	Contents	Number of Students
Class Level	Lower	First and Second Year Undergraduates	21
	Upper	Third and Fourth Year Undergraduates	17
GALT Level	Low	Scores 1-6	15
	High	Scores 9-12	16

Online exercises

Students completed online exercises 24-hours prior to taking an in-class examination. As part of the exercises, students *analyzed* sample answers to essay test questions for completeness and cogency using a 5-point Likert scale (1 = extremely complete/ cogent, 5 = extremely incomplete/ non-cogent). In some cases, students were asked to *plan* their responses to essay questions. In addition, on certain essay questions, students were asked to *construct* a response within the context of the online exercise (without access to previous screens) or as part of an in-class examination. Table 2 provides an overview of the type of activity the students were asked to complete for each question in the study.

Table 2. Components required by exercises and examinations.

Question	Topic	Student analysis of essay answers	Student plan for essay answers	Assessment of student responses
Exercise 3-1	IR Absorption	X		
Exercise 3-2	Spectra		X	
Examination 3-1	Spectra			X
Exercise 4-1	Gasoline	X		
Exercise 4-2	Entropy		X	
Exercise 4-3	Entropy			X
Examination 4-1	Gasoline			X
Exercise 5-1	Entropy no. 2			X

Student essay responses were assessed using three subscores: completeness, cogency, and achievement. Subscores were calculated by the researchers, with discrepancies discussed and resolved. Completeness and cogency were evaluated on a scale from 1 (extremely complete/cogent) to 5 (extremely incomplete/non-cogent). The achievement subscore was evaluated on a scale of 0-8 points, with points awarded for inclusion of pertinent scientific concepts/facts and logical arguments. Points were subtracted for inclusion of irrelevant scientific concepts/facts and/or weak arguments.

Results

Can students recognize a complete and cogent response to a chemistry essay question? Students were asked to rate four sample answers (labelled A-D) for each of two online essay questions. Sample answers were rated by students in terms of both completeness and cogency on a 5-point Likert scale as described before.

The analysis of the data included parametric methods such as ANOVA even though an ordinal Likert scale was used as the instrument in this study. Parametric methods of analysis were chosen over the nonparametric methods such as Chi Square on the basis of small sample size per cell, loss of power with nonparametric approaches, common use of parametric analysis methods in analyzing Likert scale data in educational literature, and doubts raised by statisticians on the absolute inappropriateness of parametric methods for ordinal data (Velleman and Wilkinson, 1993).

In this study, a mixed between-within subjects ANOVA was used to investigate the difference between student ratings of the four answers (A-D) for each of the two questions. For this test, the dependent variable was the Likert rating of the answers; the within-subjects variable was answer choice (A-D); and the between-subjects variables were GALT and undergraduate class level.

A significant main effect ($p < 0.05$) for answer choice (A-D) was found for both questions in terms of completeness and cogency. Overall, students rate the four multiple choice answers significantly differently on both completeness and cogency. Results of the ANOVA tests are provided in Table 3.

Table 3. Difference in completeness and cogency for answers A-D.

Question	Complete			Cogent		
	Wilks' Lambda	F (3,24)	p	Wilks' Lambda	F (3,24)	p
Exercise 3-1	0.44	10.08	0.00	0.34	15.56	0.00
Exercise 4-1	0.13	51.59	0.00	0.13	52.25	0.00

Post hoc testing, with an LSD adjustment, was used to locate the significant differences in student ratings of completeness and cogency for the different answers (Table 4). For exercise 3-1, answer B was evaluated by the researchers as the most complete/cogent answer. Students rated answers B and D as equally complete and cogent, as evidenced by the non-significant difference between their means. Upon review, the researchers determined that these two answers were equally complete, but answer B was more cogent, thus it is understandable that students would have trouble differentiating between the two. A significant difference was found between the means for answers B/D and those for A and C. Overall, B and D were rated significantly more complete/cogent on the Likert scale than answers A and C.

In exercise 4-1, answer C was determined by the researchers to be the most complete/cogent answer. Students also rated answer C as significantly more complete and cogent than all other answers.

Table 4. *Post hoc* testing for differences in completeness and cogency.

Question	Best answer	Complete	Cogent	Distracters	Complete		Cogent	
		Mean ^a (SE) ^b	Mean (SE)		Mean (SE)	p ^c	Mean (SE)	p
Exercise 3-1	B	2.5 (0.19)	2.4 (0.18)	A	3.7 (0.16)	0.00	3.8 (0.18)	0.00
				C	3.7 (0.22)	0.00	3.7 (0.22)	0.00
				D	3.1 (0.20)	0.09	2.9 (0.21)	0.17
Exercise 4-1	C	1.7 (0.21)	1.6 (0.20)	A	4.4 (0.12)	0.00	4.3 (0.14)	0.00
				B	3.0 (0.16)	0.00	3.1 (0.20)	0.00
				D	2.4 (0.18)	0.02	2.3 (0.18)	0.02

^a 1=Extremely complete/cogent, 5=Extremely incomplete/non-cogent

^b SE= Standard Error of the Mean

^c Significance of difference between means for Best Answer and Distracter

For these exercises, no significant main effects were found for either GALT or class level. This indicates that there is no significant difference in student ratings of each answer (A-D) based on logical reasoning ability or level of academic experience. In addition, there were no significant interaction effects between these variables.

Overall, students selected the best answers (B/D for exercise 3-1 and C for exercise 4-1) as the most complete and cogent responses. However, mean student rankings of these responses ranged from slightly complete/cogent to neutral. These results indicate that although students are able to identify the best essay response in both exercises, they are not able to accurately judge the absolute completeness/ cogency of these responses. Mean student rankings for these correct responses are given in Table 5

Table 5. Student rating of correct essay answers.

Question	Complete ^a		Cogent ^a	
	Mean	Std. Error	Mean	Std. Error
Exercise 3-1 (B)	2.5	0.19	2.4	0.18
Exercise 4-1 (C)	1.7	0.21	1.6	0.20

^a1=Extremely complete/cogent, 5=Extremely incomplete/non-cogent

Can students construct a complete and cogent response to a chemistry essay question? Student responses to four chemistry essay questions were evaluated by the researchers in terms of completeness and cogency on a 5-point Likert scale (1 = extremely complete/cogent, 5 = extremely incomplete/non-cogent). Overall, student responses were evaluated as

moderately complete ($M=2.7$, $SD=1.3$) and cogent ($M=2.9$, $SD=1.4$). Mean evaluations of student responses for each question are given in Table 6.

Table 6. Evaluation of student essay responses.

Question	Complete ^a		Cogent ^a	
	Mean	SD	Mean	SD
Examination 3-1	2.5	1.2	3.0	1.3
Exercise 4-3	2.5	1.4	2.9	1.6
Examination 4-1	3.1	1.4	3.3	1.5
Exercise 5-1	2.5	1.3	2.2	1.3
AVERAGE	2.7	1.3	2.9	1.4

^a 1=Extremely complete/cogent, 5=Extremely incomplete/non-cogent

A two-way between-subjects ANOVA was used to investigate students' ability to construct quality essay responses. GALT level and class level were used as independent variables, with student subscores (completeness, cogency, and achievement) on four chemistry essay questions as dependent variables. No significant main effect was found for either GALT or class level on the subscores of three of the four exercises (Table 7). The fourth exercise differed only in the cogency subscore of student responses, with responses from high GALT students rated significantly more cogent ($M=2.6$, $SD=1.2$) than responses from low GALT students ($M=3.5$, $SD=1.3$). In addition, responses from lower classmen (first and second year undergraduates) were rated significantly more cogent ($M=2.5$, $SD=1.2$) than responses from upper classmen (third and fourth year undergraduates) ($M=3.5$, $SD=1.2$). The statistical significance of this difference should be further investigated in a larger study. Overall, in this study, data indicate that students perform equally well in terms of completeness, cogency, and achievement regardless of GALT or class level.

Table 7. ANOVA main effects for GALT and class level on student essay responses.

Question	Independent variable	Degrees of freedom	Completeness		Cogency		Achievement	
			F	p	F	p	F	p
Examination 3-1	GALT level	1,26	1.86	0.18	6.66	0.02	2.76	0.11
	Class level		3.00	0.10	8.07	0.01	1.58	0.22
Exercise 4-3	GALT level	1,26	1.09	0.31	2.01	0.17	2.43	0.13
	Class level		0.05	0.82	0.71	0.41	1.24	0.28
Examination 4-1	GALT level	1,26	0.06	0.81	0.003	0.75	0.85	0.37
	Class level		0.73	0.40	0.74	0.40	0.01	0.94
Exercise 5-1	GALT level	1,18	1.92	0.18	1.55	0.23	1.23	0.28
	Class level		0.28	0.60	1.66	0.22	0.88	0.36

Does the practicing of evaluating sample essay responses for completeness and cogency increase student ability to construct quality answers to the same questions at a later time? Students were asked to rate the completeness and cogency of essay answers to one chemistry essay question that was later given in a testing situation. A Pearson Product-Moment Correlation Coefficient was used to investigate the relationship between students' evaluation of the correct essay answer and the completeness/cogency of the students' own answers to the same question in a testing situation. There was no significant correlation for either completeness [$r=-0.032$, $p=0.846$] or cogency [$r=0.105$, $p=0.524$]. In this study, students' ability to construct complete/cogent essay responses is therefore not correlated with their ability to rate the completeness/cogency of responses to the same question. It was hypothesized that prior evaluation of a question and its possible answers would increase the

quality of the students' responses to the same question. The data suggest that this is not necessarily true since the planning and the students' answers to the same question in a testing situation given within 24 hours showed no significant correlation.

What is the relationship between students' planned and actual responses to chemistry essay questions? For two questions (exercises 3-2 and 4-2), students were presented with a question and asked to plan a response by selecting options from the following list to include in their answer:

- *Definition* of science principles or concepts.
- *Use* of science principles or concepts.
- Graph.
- Drawing or diagram.
- Chemical equation.
- Calculation.
- Discussion at the molecular level.
- Example from the real world.
- Other.

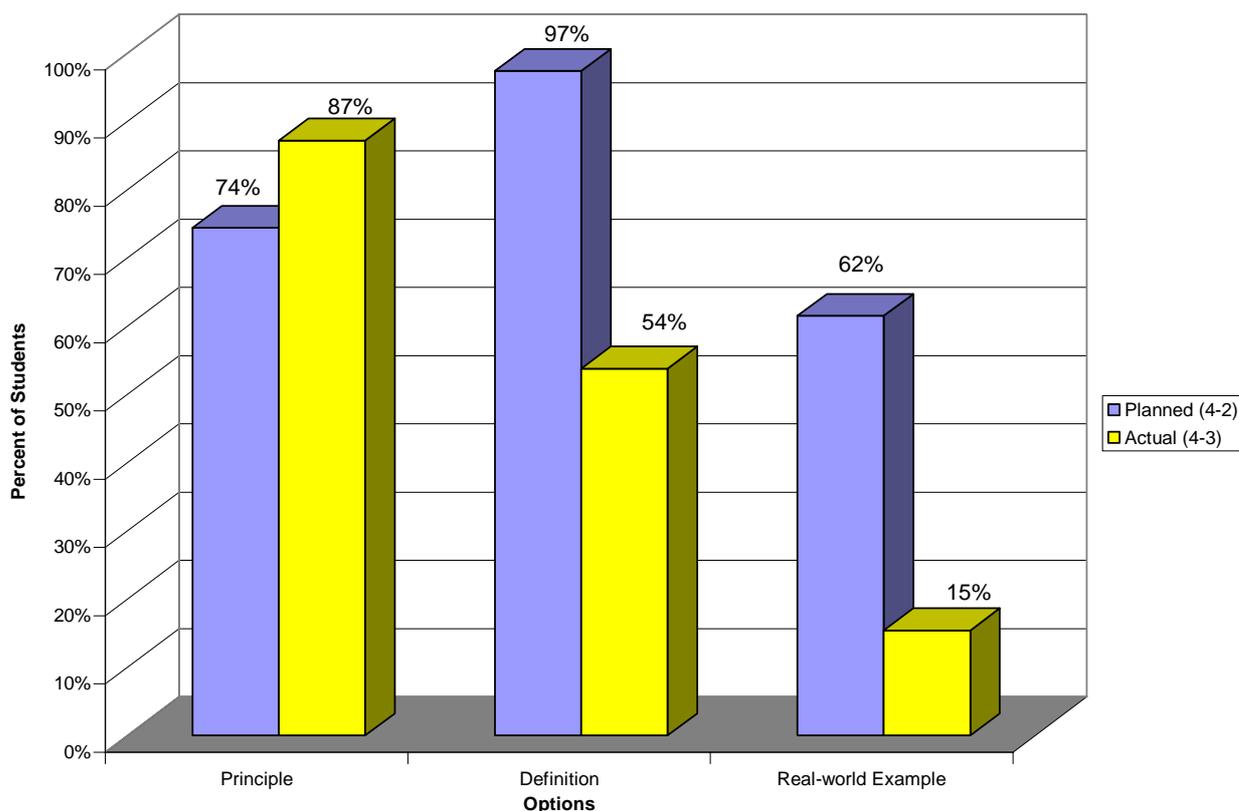
Although selecting options to include in a response is not the same as a detailed plan for constructing an essay answer, the selection of options by the students suggests an overall reflection on the type of information that would be included in an essay response.

Following these exercises, students were asked to construct an answer to the same question either online or in a testing situation. Student responses were analyzed to determine which of the criteria from the above list were included. Analysis shows that students rely primarily on principles and definitions to both plan and answer essay questions. Students' use of definitions in their responses, however, was lower than planned, while students' use of principles in their responses was higher (Figure 2). The general match between the specific options selected from the list during the planning and the evaluation of options actually used in the answer suggests that students did choose options that reflected their intention in terms of answering the question.

Discussion

Data show that students are able to identify correct essay answers, but are not able to judge accurately their absolute completeness or cogency. Similarly, when asked to construct their own responses, students are unable to provide extremely complete or cogent arguments.

One would expect students with high logical reasoning ability and/or extensive experience at the university level to be better able to construct and identify quality responses to chemistry essay questions. This expectation, based upon GALT and class level, is not confirmed in this study. This may indicate two things: 1) students are approaching the study of chemistry as novices and 2) students have difficulty applying quality essay-writing approaches in this content area. As expected of novice essay writers, students in this study rely primarily on principles and definitions to plan and answer essay questions.

Figure 2. Planned vs. actual student response for exercise 4-2 vs. 4-3.

Implications for researchers

In order to rely on essay questions as a measure of student understanding, the questions themselves must require a demonstration of understanding, analysis, or application. Simple open-ended questions that rely on recall should not be classified as essay questions that test for understanding. Although chemistry courses may never rely totally on essay questions to test for knowledge and understanding, more teachers are including essay questions as a component of their assessments. These assessments are used both to test for understanding on the part of their students and as a way to gauge the impact of teaching innovations on student learning. Research that uses student responses to chemistry essay questions as the basis for achievement is in jeopardy due to the demonstrated inability of students reliably to recognise and construct complete and cogent responses. Student exposure to sample answers with a range of completeness and cogency was insufficient training to improve student success. The literature supports this finding, and further suggests that training in essay-writing in chemistry must be deliberate and extensive (Russell, 2004). Thus, researchers who choose to rely on student responses to essay questions as an indication of student achievement or conceptual understanding should be aware of the intervening effect of inadequate student essay-writing ability.

References

Abraham M.R., Gryzbowski E.B., Renner, J.W. and Marek E.A., (1992), Understandings and misunderstandings of eighth graders of five chemistry concepts found in textbooks, *Journal of Research in Science Teaching*, **29**, 105-120.

- Bodner G., (1991), I have found you an argument: the conceptual knowledge of beginning chemistry graduate students, *Journal of Chemical Education*, **68**, 385-388.
- Cavallo A.M.L., McNeely J.C. and Marek E.A., (2003), Eliciting students' understandings of chemical reactions using two forms of essay questions during a learning cycle, *International Journal of Science Education*, **25**, 583-603.
- Danili E. and Reid N., (2005), Assessment formats: do they make a difference?, *Chemistry Education Research and Practice*, **6**, 204-212.
- Kovac J. and Sherwood D.W., (1999), Writing in chemistry: an effective learning tool, *Journal of Chemical Education*, **76**, 1399-1403.
- Moore J.W., (1997), Assessment, achievement, and understanding, *Journal of Chemical Education*, **74**, 477.
- Novak, J.D. and Gowin, D.B., (1984), *Learning how to learn*, New York, Cambridge University Press.
- Oliver-Hoyo M.T., Allen D., Hunt W.F., Hutson J. and Pitts A., (2004), Effects of an active learning environment: teaching innovations at a research I institution, *Journal of Chemical Education*, **81**, 441-448.
- Roadrangka V., Yeany R.H. and Padilla M.J., (1982), *Group assessment of logical thinking*, University of Georgia.
- Russell A.A., (2004), Calibrated peer review: a writing and critical-thinking instructional tool. In S. Cunningham and Y.S. George (Eds.), *Invention and impact: building excellence in undergraduate science, technology, engineering and mathematics (STEM) education*. Washington, DC, American Association for the Advancement of Science.
- Sweller J., (1994), Cognitive load theory, learning difficulty and instructional design, *Learning and Instruction*, **4**, 295-312.
- Velleman P.F. and Wilkinson, L., (1993), Nominal, ordinal, interval, and ratio typologies are misleading, *The American Statistician*, **47**, 65-72.
- Wandese J.H., Mintzes J. and Novak J.D., (1994), Research on alternative conceptions in science, in D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177-210), New York, Macmillan.