EJPCE
Eurasian Journal of Physics and Chemistry Education

Eurasian J. Phys. & Chem. Educ. 4(2): 157-172, 2012

journal homepage: http://www.eurasianjournals.com/index.php/ejpce

Test Item Formats in Finnish Chemistry Matriculation Examinations

Greta Tikkanen and Maija Aksela*

Department of Chemistry, Chemistry Teacher Education Unit, University of Helsinki, Finland

Received: 20 February 2012 - Revised: 07 May 2012 – Accepted: 11 May 2012

Abstract

Summative assessment plays an essential role in the chemistry education. This paper presents an analysis of Finnish chemistry matriculation examination questions according to test item format, and some examples of the analysis and examination questions. The research data consisted of 257 chemistry questions from 28 matriculation examinations between 1996 and 2009. Qualitative approach and theory-driven content analysis method were employed in the research. This research was guided by the following question: What kinds of test item formats are used in chemistry matriculation examinations? The research indicates that summative assessment was used diversely in chemistry matriculation examinations. The tests included various test item formats, and their combinations. The majority of the test questions were constructed-response items that were either verbal, quantitative, or laboratory-related items, symbol items, or combinations of the aforementioned. The studied chemistry matriculation examinations seldom included selected-response items that can be either multiple-choice, binary-choice, or matching items. The classification framework developed in the research can be applied in chemistry and science education, and also in educational research.

Keywords: **C**hemistry matriculation examination questions, Classification, Constructed-response items, Selected-response items, Test item formats

Introduction

Assessment lies at the heart of the chemistry education. Teachers teach and students study towards success on tests (Tamir, 2003). Thus, assessment points to what is considered relevant and ignores what is perceived to be unimportant (Doran, Lawrenz & Helgeson, 1994). Assessment in chemistry education can be divided into three main types: diagnostic, formative and summative assessment (Black, 2004; Doran et al., 1994; Harlen, 2004). Student assessment is often based on summative assessment that is both predictive and comparable, and also gives an overview of students' previous learning obtained during an instructional unit (Black, 2004).

In general, summative assessment is implemented at the end of an instructional unit to measure and document student achievement in proportion to other students' performance, or some predetermined instructional standards (Doran et al., 1994; McMillan, 2008). Various tests and examinations are typical summative assessment tools (McMillan, 2008).

Summative test consists of test items that represent different item formats (Haladyna, 2004). All item formats have their advantages and limitations (Uusikylä & Atjonen, 2005). They also have a didactic function because they reflect what is considered important (Salmio, 2004). For example, the test item formats used in matriculation examinations have an impact on Finnish high-school education (Lindblom-Ylänne, 2003).

*Corresponding Author, Phone: +358-50-5141450; Fax+358-9-1915 0466 E-mail: maija.aksela@helsinki.fi

ISSN: 1306-3049,

Test items can be classified in different ways (Welch, 2006). They are often divided into objective and subjective test items according to the grading system (Rodriquez, 2002). Another commonly used classification method is based on the answering system of the test items. It's partly parallel with the aforementioned classification (Rodriquez, 2002), and the test items are divided into two main categories: *selected-response items* and *constructed-response items* (Hancock, 2007; Hogan & Murphy, 2007; McTighe & Ferrara, 1998; Osterlind, 1998; Popham, 2003; Rodriquez, 2002). The terms *supply item*, *free-response item*, or *open-response item* may also be used when referred to constructed-response items (Rodriquez, 2002).

McTighe and Ferrara (1998) divided test items into two main categories: selected-response items (*multiple-choice*, *true-false*, *matching*, and *enhanced multiple-choice items*) and constructed-response items (*brief constructed-response items*, *performance-based assessment*). Martinez (1999) divided test items into multiple-choice items and constructed-response items (*discrete* and *extended-performance constructed-response items*).

Table 1. A summary of selected-response and constructed-response items*

SELECTED-RESPONSE ITEMS	**CONSTRUCTED-RESPONSE ITEMS**
Definition:	*Definition*:
Selected-response items require students to **select** the answer from the given options.	Constructed-response items require students to **construct** the response by themselves.
Advantages:	*Advantages*:
+ excellently applicable for measuring the learning of broad areas of knowledge	+ excellently applicable for assessing students' concrete skills and products
+ grading process is objective and fast	+ may give better information on students' understanding and learning compared to selected-response items
+ answering process is fast	+ applicable for assessing individuality, creativity and originality
+ excellent coverage of the learning area	+ possibility of guessing is minimal
+ applicable for measuring broad learning at several different cognitive levels	+ creating constructed-response items is usually easy
Limitations:	+ may resemble real-life problems
– not applicable for measuring students' concrete skills and products	*Limitations*:
– difficult to measure creativity and critical thinking	– grading process is usually subjective, arduous, challenging and expensive, and it always requires developing of a consistent grading model
– possibility of guessing	– answering process is usually time-consuming
– usually differ significantly from real-life problems	– covering broad areas of learning effectively is challenging
– creating of high-quality selected-response items is challenging	*Sub categories*:
Sub categories:	▪ Short-answer items
▪ Multiple-choice items	▪ Performance assessments
▪ Binary-choice items	
▪ Matching items	

*e.g: Downing, 2002, 2003, 2006; Hogan & Murphy, 2007; Holt & Kysilka, 2005; Leuenberger, 2001; McTighe & Ferrara, 1998; Pelton & Pelton, 2006; Plake, 2005; Popham, 2003; Quellmalz & Hoskyn, 1997; Scheerens et al., 2003; Stiggins & Erter, 2004.

Hogan and Murphy (2007) divided test items into selected-response items (*multiple-choice, true-false,* and *matching items*) and constructed-response items (*performance assessments, portfolios*). A summary of selected-response and constructed-response items is shown in Table 1.

Selected-response items are widely used in chemistry assessment. They include a stem (e.g. question, problem) and a series of response alternatives (Hogan & Murphy, 2007; McTighe & Ferrara, 1998). The items require students to select the answer from the given alternatives (Downing, 2006; Kraska, 2008; Hogan & Murphy, 2007; McTighe & Ferrara, 1998; Osterlind, 1998).

Selected-response items can be used effectively to get broad information on students' chemistry learning (Downing, 2006; McTighe & Ferrara, 1998). They are fast to answer, so it's possible to include several items in each test. Therefore, it's possible to cover a broad area of learning, such as high school chemistry contents, comprehensively. (McTighe & Ferrara, 1998; Stiggins & Erter, 2004)

The grading process of selected-response items is fast and objective and the items are generally machine-scorable (Downing, 2006). This is a significant advantage, especially in large-scale assessments such as matriculation examinations.

Selected-response items are particularly useful in measuring the learning of broad areas of knowledge at a wide range of cognitive levels (Downing, 2006). They provide broad information on students' learning of factual knowledge (Holt & Kysilka, 2005; Stiggins & Erter, 2004; Quellmalz & Hoskyn, 1997), concepts and principles, and application of basic skills (McTighe & Ferrara, 1998). Selected-response items can also be used in assessing students' higher-order cognitive skills.

Selected-response items have limitations. They're not applicable for assessing concrete skills or products (Downing, 2006; Stiggins & Erter, 2004). It's also very challenging to measure students' critical thinking and creativity with them. Selected-response items require students to identify the correct answer from the given alternatives, so they differ significantly from the real- life problems that usually have several correct answers. Therefore, their excessive use can lead to a distortion of students' understanding of the nature of knowledge and learning. (McTighe & Ferrara, 1998) There's always a possibility of guessing related to using of selected-response items as an assessment tool. However, its impact on the overall assessment results is usually not significant (Downing, 2003).

The very demanding creating process is one of the biggest challenges pertaining to the use of selected-response items in assessment (Downing, 2006). Creating of high-quality items is a time-consuming process that requires good resources and expertise (Downing, 2006; Pelton & Pelton, 2006).

Multiple-choice items are the best known and most widely used form of selected-response items (Hogan & Murphy, 2007; McTighe & Ferrara, 1998). They are versatile items, which are applicable for assessment of complex learning outcomes better than other types of selected-response items (Miller et al., 2008). A typical multiple-choice item contains a stem and several response alternatives (Haladyna, 2004; Nitko & Brookhart, 2007; Popham, 2003; Wakeford, 2003). Students must select the correct or best answer from the given options (Haladyna, 2004).

Binary-choice items and matching items are commonly used subtypes of selected-response items (Kraska, 2008; Quellmalz & Hoskyn, 1997), although they may also be considered as variations of multiple-choice items (Downing, 2006; Hogan & Murphy, 2007). Binary-choice items (e.g. *true-false items*) require students to select the correct answer from two response alternatives (Haladyna, 2004). It's typically easier to create binary-choice items than multiple-choice items (Scheerens et al., 2003), and they're also very time-efficient items

(Haladyna, 2004; Woolfolk, 2007). However, the possibility of guessing is more significant in binary-choice items compared to multiple-choice items because there're only two response alternatives (Haladyna, 2004; Miller et al., 2008). One of the major limitations pertaining to binary-choice items is the fact that they often require students to only recognize the wrong answer instead of knowing the correct answer (Miller et al., 2008).

Matching items typically contain two lists, which contain the premises and possible answers of the item. The purpose is to match the items of these two lists according to the criteria described in the stem of the item. (Nitko & Brookhart, 2007) Matching item can be regarded as an effective series of multiple-choice items (McMahon et al., 2006), in which each premise forms a separate item (Nitko & Brookhart, 2007). Thus, they are very compact tasks (Haladyna, 2004; Nitko & Brookhart, 2007).

Matching items are usually easier to create than the regular multiple-choice items (Haladyna, 2004; Scheerens et al., 2003). On the other hand, it's required that the assessed learning area contains adequately homogeneous material that can be used as the basis of the premises and answers (Miller et al., 2008). One other limitation of the matching items when comparing with other selected-response items is the possibility to eliminate the response alternatives during the solution process (Woolfolk, 2007).

Constructed-response items are widely used in the chemistry assessment. They are test items that require students to construct the response by themselves instead of selecting it from the given options (Bennett, 1993; Downing, 2002; Hancock, 2007; Hogan & Murphy, 2007; Osterlind, 1998). The category of constructed-response items is very broad, and it includes a variety of tasks (Bennett, 1993; Martinez, 1999). The chemistry constructed-response items may, for example, require students to make calculations, construct graphic presentations or extended essays, or write and balance chemical equations (Tarendash, 2006).

Constructed-response items have many advantages. They are usually easier to create than multiple-choice items (Downing, 2002). The impact of the possibility of guessing on the assessment results is also insignificant (Plake, 2005). Constructing the response is clearly a more demanding and authentic task for the students than simply recognizing the correct answer from the given alternatives (Downing, 2002; Popham, 2003). Usually, students need to use both conceptual and strategic knowledge when formulating the response instead of simply memorizing the facts (Holt & Kysilka, 2005).

Constructed-response items may possibly give better information on students' understanding and learning than selected-response items (Leuenberger, 2001). They can also be used in assessing students' individuality and originality, and their ability to apply knowledge and skills (McTighe and Ferrara, 1998). Constructed-response items also have limitations. The grading process is usually challenging, time-consuming, inefficient, and expensive (Downing, 2002), and it always requires developing of a consistent and explicit grading model (Scheerens et al., 2003). Complex constructed-response items may be particularly arduous to assess (Popham, 2003). On the other hand, it's difficult to cover broad areas of knowledge comprehensively with constructed-response items due to the long response time (Downing, 2002; Scheerens et al., 2003).

Constructed-response items may be classified in different ways (cf. Hogan & Murphy, 2007; Martinez, 1999; McTighe & Ferrara, 1998). Some typical sub categories of the chemistry constructed-response items are represented in the following. Short-answer items can be defined as questions that require a limited (max 1 page) written response (Wakeford, 2003). CUSE (1997) define short-answer items as questions, which require a response of one or two sentences or a short paragraph at most. Short-answer items may also be defined as questions, which require a response of a word, sentence, number, or a symbol (Miller et al., 2008; Nitko & Brookhart, 2007).

Short-answer items may be used to obtain information on the learning of terminology, facts, symbols, principles, classifications, and methods (Miller et al., 2008; Nitko & Brookhart, 2007). On the other hand, they can be used to measure students' ability to make simple interpretations of numerical or graphical data (Miller et al., 2008). Short-answer items may be science problems, and they may require students to manipulate symbols, or balance chemical equations (Nitko & Brookhart, 2007).

Short-answer items are usually fairly easy and quick to create and grade (Wakeford, 2003). On the other hand, the questions are quite fast to answer, so several items can be included in each test. Therefore, a broad learning area can be covered with the items. (Popham, 2003) The impact of guessing on the assessment results is also quite insignificant (Miller et al., 2008; Nitko & Brookhart, 2007). Short-answer items have limitations. They are often ambiguous, which makes it difficult to grade them objectively (Nitko & Brookhart, 2007; Popham, 2003). Creating of the items that measure higher-order cognitive levels is also challenging (Wakeford, 2003). Therefore, short-answer items are not particularly well suited for measuring complex learning (Woolfolk, 2007), but they are often used to assess memorization of factual knowledge (Miller et al., 2008; Popham, 2003).

Short-answer items may, however, give also information on students' deeper understanding and problem-solving skills. The questions requiring interpretation of diagrams, graphs, charts, and images are examples of more demanding short-answer items. (Miller et al., 2008)

Essay is a well-established summative test item in chemistry assessment. It requires students to construct an extended written response to a question or problem (Brooks & Crippen, 2006; Nitko & Brookhart, 2007; Wakeford, 2003).

Essay items have many advantages. They're fairly easy and fast to create (Wakeford, 2003). Essays are also applicable for assessing students' higher-order cognitive skills (Wakeford, 2003), thinking processes, and creativity (Woolfolk, 2007). They may give information on students' understanding and ability to apply knowledge in novel situations (CUSE, 1997).

Essay items have limitations. The grading system is fairly subjective (Heinonen & Viljanen, 1980; Miller et al., 2008; Woolfolk, 2007), and student's linguistic talent and handwriting may have an impact on the assessment (Heinonen & Viljanen, 1980). Essay items are also very arduous to grade (Woolfolk, 2007). One of the biggest limitations of essays is their lack of coverage because each test may include only a couple of essay items at most. Therefore, they are not suitable for measuring an overall learning of broad areas of learning. (Wakeford, 2003; Miller et al., 2008; Nitko & Brookhart, 2007).

Quantitative problems are widely used in chemistry assessment. Most of the questions are closed and well-defined routine exercises that typically have only one correct solution and include all the information needed in the solution process (Reid & Yang, 2002). Generally, students only need to apply simple and familiar algorithms (Bennett, 2008). However, real problems that emphasize conceptual understanding and application skills should also be used in the chemistry assessment (Haláková & Prokša, 2007; Phelps, 1996). The problems should also be contextually meaningful and related to students' everyday life (Murphy & McCormick, 2006).

Laboratory experiments play a key role in chemistry education and assessment (Doran et al., 1994; Lunetta et al., 2007). The process skills pertaining to laboratory work can be assessed both with written and practical performance tasks. Written tasks may require constructing of different kinds of products such as research reports (Ferrer, 2008). They may also require students to design experimental methods (Huffman, 2002), or construct and interpret graphic presentations (Temiz et al., 2006). Process skills can also be assessed with

multiple-choice items. For example, they may include a research question on the basis of which students must select the most suitable research method. (Huffman, 2002)

Practical performance tasks have several advantages. They may give different kind of information on students' chemistry learning than written tests (Doran et al., 1994). On the other hand, practical tasks may increase students' interest towards chemistry, improve their understanding of the nature of chemistry, and support the construction of the conceptual and procedural knowledge of chemistry (Lunetta et al., 2007).

Practical tasks are particularly suitable for assessing students' process and problem-solving skills (Doran et al., 1994; Huffman, 2002). Including them in the chemistry assessment may also increase the use of experimental methods in chemistry education (Tamir, 2003; Temiz et al., 2006). Practical tasks have limitations. They're often expensive, time-consuming and difficult to create, and it's challenging to guarantee the high reliability and validity of the assessment. (Doran et al., 1994; Tamir, 2003)

The matriculation examination is the dominant summative assessment tool in Finnish high schools. It consists of at least four tests in different subjects. Chemistry test is one of the optional tests. The matriculation examination is a very traditional institution that has a great impact on both teaching and learning of chemistry in Finnish high schools (Aksela & Juvonen 1999). Therefore, it is very important to find a detailed answer to the following question: *What kinds of test item formats are used in chemistry matriculation examinations?*

Methodology

Chemistry matriculation examination questions are classified according to the item format in this research. The research data consisted of 257 chemistry questions from 28 matriculation examinations between 1996 and 2009. Qualitative approach and theory-driven content analysis method were employed in the research (Cohen et al., 2007).

There are several phases in the research. In the first phase, chemistry matriculation examination questions are classified into two main categories: 1) the test items that include only one item format, and 2) the test items that include several item formats. The test items of category 1 are then classified into sub items according to the classification framework shown in Figure 1.

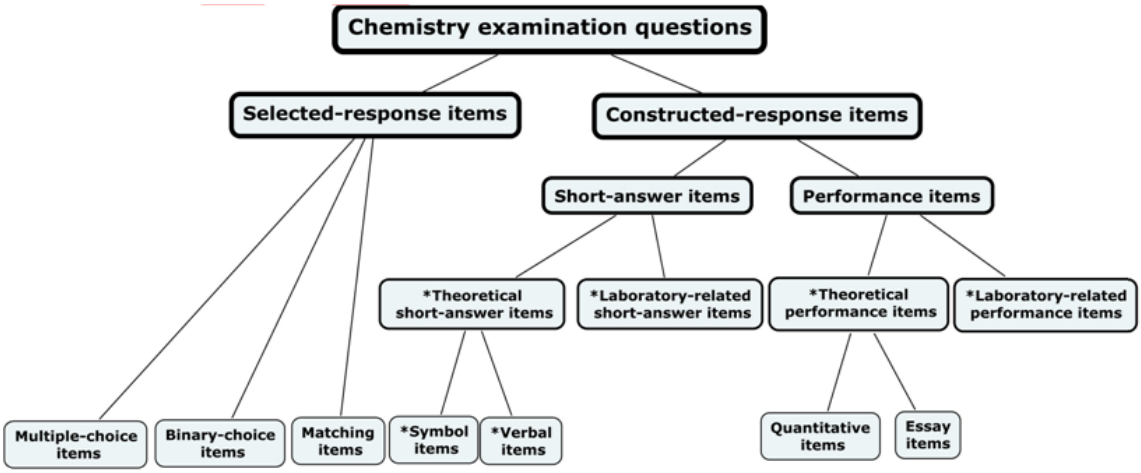


Figure 1. The classification framework of the research.

The classification framework has been constructed on the basis of what is discussed in the research literature. The analyzed research data has also been taken into account when

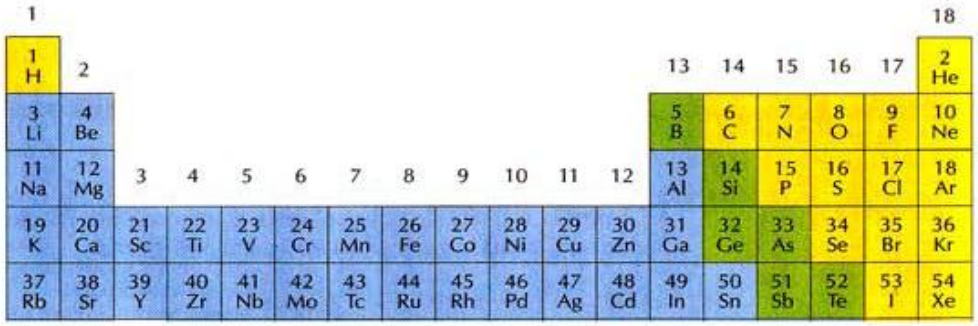
creating the framework. The categories formed on the basis of the research data have been marked with asterisk (*) (see Figure 1). The definitions of the test items employed in the research are shown in Table 2.

Table 2. The test item definitions employed in the research.

SELECTED-RESPONSE ITEMS ▪ Selected-response items require students to select the answer from the given alternatives.	**Multiple-choice items** ▪ Multiple-choice items require students to select the answer from at least three given alternatives. **Binary-choice items** ▪ Binary-choice items require student to select the answer from two given alternatives. **Matching items** ▪ Matching items require students to match premises and possible answers according to the criteria described in the stem of the item.
CONSTRUCTED-RESPONSE ITEMS ▪ Constructed-response items require students to construct the response by themselves. ▪ They may also require concrete performances.	**Symbol items** ▪ Symbol items are theoretical short-answer items that are not related to laboratory work, and are answered in symbols. **Verbal items** ▪ Verbal items are theoretical short-answer items that are not related to laboratory work, and are answered in words or a few sentences. **Laboratory-related short-answer items** ▪ Laboratory-related short-answer items measure students' knowledge and/or skills related to practical laboratory work. ▪ They may include theoretical symbol or verbal sub items. ▪ They require students to plan or describe an experiment or laboratory method, and/or measure their knowledge of practical laboratory work. **Quantitative items** ▪ Quantitative items are theoretical performance items that are not related to laboratory work, and require students to make calculations. **Essay items** ▪ Essay items are theoretical performance items that are not related to laboratory work, and require students to construct an extended written response to a question or problem. **Laboratory-related performance items** ▪ Laboratory-related performance items measure students' knowledge and/or skills related to practical laboratory work. ▪ They may include theoretical sub items. ▪ They require students to plan, describe, explain, or conduct experiments or laboratory methods, and/or require them to interpret or manipulate experimental data quantitatively, or construct graphic representations based on experimental data. ▪ These items may also measure students' knowledge of practical laboratory work.

The chemistry examination questions that include only one item format are classified into two main categories: selected-response and constructed-response items, which are then classified into sub items according to the classification framework. Examples of the classification process are shown in Table 3.

Table 3. Examples of the classification process of the examination questions that contain only one item format.

Question 1. (Verbal item) Explain the following concepts: a) element b) chemical compound, c) molecule, d) zwitter ion, e) azeotrope, f) rasemic mixture. Classification process: constructed-response item → short-answer item → theoretical short-answer item → verbal item
Question 2. (Symbol item) Construct the structural formula for the following compounds: a) 2-penthanone, b) ethylmethylamine, c) cyclohexanol, d) 1,3-butadiene, e) 1,4-dihydroxybenzene, f) trans-2,3-dichloro-2-buthen. Classification process: constructed-response item → short-answer item → theoretical short-answer item → symbol item
Question 3. (Multiple-choice item) See below for the periodic table and select from it a) a semiconductor b) an element that shows allotropy c) an element, which is an alkaline oxide d) the first transition element e) the most powerful oxidant f) the elements whose density is lower than the density of air (NTP).  Classification process: selected-response item → multiple choice item
Question 4. (Essay item) The isomerism of organic compounds. Classification process: constructed-response item → performance item → theoretical performance item → essay item
Question 5. (Laboratory-related performance item) An aqueous solution contains an unknown amount of hydrochloride. Design an experimental method which can be used to find out the amount of hydrochloride in solution accurately. Describe the method in detail, its chemical basis, and the necessary laboratory arrangements. Classification process: constructed-response item → performance item → laboratory-related performance item

In the next phase of the research, the chemistry examination questions that include several item formats are classified into sub items according to the classification framework (see Figure 1). Examples of the classification process are shown in Table 4**.**

Table 4. Examples of the classification process of the examination questions that contain several item formats.

Question 1 (Matching item + Symbol item)

Listed below are a number of compounds and some chemical concepts. Which of these can be combined with each other? Construct the structural formula for each compound, and represent all the related concepts by using the given numbering system.

a) ethyne	1) planar structure
b)1,1- dichloroethene	2) *cis-trans*-isomerism
c) benzene	3) polar molecule
d) pyridine	4) optical isomerism
e) 2-butanol	5) linear (rod-shaped) structure
f) 2,2-dimethylpropane	6) aromatic hydrocarbon
	7) heterocyclic compound
	8) tetraedric carbon atom

Explanation: This item is a matching item because it requires students to match compounds and chemical concepts. It's also a symbol item because students have to write structural formula.

Question 2 (Quantitative item + Verbal item)

Equilibrium occurs in the saturated aqueous solution of calcium fluoride

$$CaF_2(s) \rightleftharpoons Ca^{2+}(aq) + 2F^-(aq); K_L = 3{,}9 \cdot 10^{-11}(mol/l)^3$$

a) Calculate the fluoride ion concentration of a saturated aqueous solution of calcium fluoride.

b) How does the fluoride ion concentration change when a small amount of solid calcium chloride is added to the saturated aqueous solution of calcium fluoride? Explain your answer.

c) How does the solubility of a salt change when a small amount of hydrochloride is added to the saturated aqueous solution of calcium fluoride? Explain your answer.

Explanation: This item is a quantitative item because it requires students to calculate in the sub item a). It's also a verbal item because the sub items b) and c) are answered verbally.

Question 3 (Quantitative item + Symbol item)

An organic compound contains 34,3 mass % of carbon, 6,7 mass % of hydrogen, 13,3 mass % of nitrogen, and 45,7 mass % of oxygen.

a) Find the empirical formula (ratio formula) for the compound.

b) What is the molecular formula for the compound when its relative molecular mass is 105?

c) Construct a possible structural formula for the compound when we know it's a natural amino acid.

Explanation: This item is a quantitative item because it requires students to calculate in the sub items a) and b). It's also a symbol item because students have to construct a structural formula in the sub item c).

Peer review was used to guarantee the reliability of the results. 10% of the examination questions were picked at random and analyzed by a research scientist specialized in the field of chemistry education. The value of Cohen's kappa coefficient was calculated based on the peer review results.

The Cohen's kappa value for the classification was 0,877. The high kappa value indicates an excellent inter-rater agreement between the raters, and thus high reliability for the research.

Results

The majority (62%) of the chemistry matriculation examination questions included only one item format in 1996–2009. The rest of the questions (38%) were different kinds of combination items. Almost all the analyzed questions were either constructed-response items, or combination items that required at least one constructed-response sub item. The distribution of the examination questions is shown in Table 5. The percentages are rounded to the nearest whole percent.

Table 5. Distribution of the chemistry matriculation examination questions (N=257) in 1996–2009.

<table>
<tr><td rowspan="4">SELECTED -RESPONSE ITEMS - 2
(1%)</td><td colspan="3">Multiple-choice items
1
(<1 %)</td></tr>
<tr><td colspan="3">Binary-choice items
1
(<1 %)</td></tr>
<tr><td colspan="3">Matching items
0
(0 %)</td></tr>
<tr><td rowspan="5">CONSTRUCTED - RESPONSE ITEMS -158
(61 %)</td><td rowspan="3">Short-answer items - 58
(23 %)</td><td rowspan="2">Theoretical short-answer items - 50
(19 %)</td><td>Symbol items - 15
(6 %)</td></tr>
<tr><td>Verbal items - 35
(14 %)</td></tr>
<tr><td colspan="2">Laboratory-related short-answer items - 8
(3 %)</td></tr>
<tr><td rowspan="2">Performance items - 100
(39 %)</td><td rowspan="2">Theoretical performance items- 78
(31 %)</td><td>Quantitative items - 38
(15 %)</td></tr>
<tr><td>Essay items 40
(16 %)</td></tr>
<tr><td colspan="3">Laboratory-related performance items - 22
(9 %)</td></tr>
<tr><td colspan="4">COMBINATION ITEMS - 97
(38 %)</td></tr>
</table>

The majority (99%) of the chemistry matriculation examination questions that contained only one item format were constructed-response items in 1996–2009. There were only two selected-response items, one multiple-choice item and one binary-choice item, included in the analyzed questions. Both items pertained to elements and compounds.

Most of the constructed-response items were theoretical performance items. They required students to construct extended written responses, or solve quantitative chemistry problems. Most essay items included information or guiding questions to help students formulate the response. Some essay items also required interpretation of graphic representations. The quantitative items were typically stoichiometric problems, although several items pertained to chemical equilibrium.

The chemistry matriculation examination questions contained a significant proportion (19%) of theoretical short-answer items. They required students to construct short verbal or symbolic responses. Most verbal items required students to explain or compare chemical concepts, or give a chemical explanation of different kinds of phenomena. Symbol items typically required writing of chemical equations, or structural formula of organic compounds.

Laboratory-related constructed-response items were also included in the analyzed examination questions. Most of them were laboratory-related performance items requiring students to plan, describe, or explain a chemical method or experiment. Some of the items

measured students' knowledge of practical laboratory work, and/or their ability to interpret or quantitatively manipulate experimental data. Constructing of graphic representations on the basis of experimental data was also required in some laboratory-related performance items.

Some laboratory-related short-answer items were also found when analyzing the chemistry matriculation examinations in 1996–2009. They mainly required students to describe or explain chemical methods or experiments in a few sentences. Students' knowledge of laboratory safety was also often measured with these items.

Table 6. Distribution of the chemistry matriculation examination questions that include several item formats, and the proportions (%) of different item combinations of all the analyzed items (N=257) in 1996–2009.

COMBINATION	**PROPORTION**
VI + SI	78 **(30 %)**
VI + QI	
VI + QI + SI	
VI + QI + BCI	
VI + BCI	
VI + MCI + SI	
VI + MCI	
VI + QI + MCI	
VI + LSAI	
VI + MI	
VI + SI + MCI + BCI	
VI + EI + SI	
VI + EI	
VI + LPI	
VI + QI + SI + BCI	
SI + QI	19 **(7 %)**
SI + MI	
SI + MCI	
SI + EI	
SI + BCI + QI	
BCI + QI	
BCI + LSAI	
LPI + MCI	
EI + LSAI	
TOTAL	97 **(38 %)**

MCI = multiple-choice item; BCI = binary-choice item; MI = matching item; SI = symbol item;
VI = verbal item; LSAI = laboratory-related short-answer item; QI = quantitative item; EI = essay item;
LPI = laboratory-related performance item

38% of the chemistry matriculation examination questions contained at least two different item formats in 1996–2009. In total, 24 different item combinations were found. Most of the questions included a verbal sub item. Therefore, it's meaningful to divide the combination items into two main categories: combination items that contain a verbal sub item and other combination items. The distribution of the combination items is shown in Table 6. The item formats are marked with abbreviations that are explained below the table.

Conclusions and Discussion

The research indicates that summative assessment was used diversely in the chemistry matriculation examinations in 1996–2009. The tests included various test item formats, and their combinations. The majority of the examination questions (99%) were either constructed-response items, or combination items that contained at least one constructed-response sub item.

The studied chemistry matriculation examinations included only two selected-response items that were multiple-choice, and binary-choice items. This result can be considered as very uncommon due to the strong position of selected-response items both in the research literature (e.g. Hogan & Murphy, 2007; Martinez, 1999; McTighe & Ferrara, 1998), and in other final high school chemistry examinations such as *International Baccalaureate* (*IB*). Therefore, this research shows that the main test item formats used in the Finnish chemistry matriculation examinations are not aligned with the test item classifications represented in the research literature.

The majority (62%) of the chemistry matriculation examination questions included only one item format. The rest of the questions were different kinds of combination items. This result shows one distinctive characteristic of the Finnish chemistry matriculation examinations because combination items are not truly discussed in the research literature.

Most of the test items that contained only one item format were items in which students were required to construct an extended or short written response. The combination items also very often included a verbal sub item. A great proportion of quantitative items were also found in the chemistry matriculation examinations. Verbal and quantitative items are widely discussed in the research literature (CUSE, 1997; Nitko & Brookhart, 2007; Reid & Yang, 2002; Wakeford, 2003). Therefore, the Finnish chemistry matriculation examinations are well aligned with the research literature when regarding these specific test item formats.

The research indicates that the experimental nature of chemistry has been taken into account when creating the chemistry matriculation examinations. The examinations contained several items, which were often similar to the verbal sections of the practical tasks described in the research literature (Huffman, 2002). For example, in many items students were required to plan experiments or manipulate experimental data.

The classification framework developed in the research (see Figure 1) can be considered as one of its main results. The main categories of the framework have been constructed on the basis of the research literature (Hogan & Murphy, 2007; Rodriquez, 2002). The sub categories of the framework such as symbol item and laboratory-related short-answer or performance item categories have been formed to make the framework applicable for test item analysis in the context of chemistry and science education. The classification framework can also be applied in educational research.

A very interesting aspect of future research in this area is a study of alignment between test item format, and test item's cognitive complexity. Another considerable research idea is an analysis of some other chemistry/science examination questions, or chemistry/science school book/classroom exercises with the classification framework developed in this research.

References

Aksela, M. & Juvonen, R. (1999). *Kemian opetus tänään* [Chemistry Education Today]. Opetushallitus [Ministry of Education]. Helsinki: Edita Oy, 38–41.

Bennett, R. E. (1993). On the meanings of constructed response. In R. E. Bennett & W. C. Ward (eds.), *Construction versus choice in cognitive measurement*. New Jersey: Lawrence Erlbaum Associates, Inc, 1–2.

Bennett, S.W. (2008). Problem solving: Can anybody do it?. *Chemistry Education Research and Practice*, 9, 60–64.

Black, P. (2004). Purposes for assessment. In J. Gilbert (ed.), *The routledgefalmer reader in science education*. London: Routledge, 189–198.

Brooks, D.W. & Crippen, K.J. (2006). Web-based practice and assessment systems in science. In J. J. Mintzes & W. H. Leonard (eds.), *Handbook of college science teaching: Theory, research, and practice*. Arlington, Va.: NSTA Press, 253.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education*. 6th edition. London: Routledge.

CUSE (Committee on Undergraduate Science Education, National Research Council), (1997). *Science teaching reconsidered*: A handbook. Washington D.C.: National Academy Press, 39–45.

Doran, R.L., Lawrenz, F. & Helgeson, S. (1994). Research on assessment in science. In D. L. Gabel, (ed.), *Handbook of research on science teaching and learning*. New York: Macmillan Publishing Company, 388–427.

Downing, S.M. (2002). Assessment of knowledge with written test forms. In G. R. Norman, C. Vleuten & C. V. D. Newble (eds.), *International handbook of research in medical education*. Dordrecht: Springer, 647–670.

Downing, S.M. (2003). Guessing on selected-response examinations. *Medical Education*, 37(8), 670–671.

Downing, S.M. (2006). Selected-response item formats in test development. In S. M. Downing & T. M. Haladyna (eds.), *Handbook of test development*. London: Routledge, 287–302.

Ferrer, L. (2008). *Building effective strategies for teaching of science'2008 ed.*. Manila: Rex Bookstore, Inc., 193–198.

Haladyna, T.M. (2004). *Developing and validating multiple-choice test items*. 3rd edition. Mahwah, NJ: Routledge, 3–4, 67–98.

Haláková, Z. & Prokša, M. (2007). Two kinds of conceptual problems in chemistry teaching. *Journal of Chemical Education*, 84(1), 172–174.

Hancock, D.R. (2007). Effects of performance assessment on the achievement and motivation of graduate students. *Active Learning in Higher Education*, 8(3), 219–231.

Harlen, W. (2004). *Teaching, learning and assessing science 5–12*. 3rd edition. London: Paul Chapman Publishing Ltd, 108–121.

Heinonen, V. & Viljanen, E. (1980). *Evaluaatio koulussa*. [Evaluation in school] Keuruu: Otava, 11–266.

Hogan, T.P. & Murphy, G. (2007). Recommendations for preparing and scoring constructed-response items: What the experts say?. *Applied Measurement in Education*, 20(4), 427–441.

Holt, L.C. & Kysilka, M.L. (2005). *Instructional patterns: Strategies for maximizing student learning*. Thousand Oaks, CA: SAGE, 116.

Huffman, D. (2002). Evaluating science inquiry: A mixed-method approach. In J. W. Altschuld & D. D. Kumar (eds.), *Evaluation of science and technology education at the dawn of a new millennium*. New York: Springer, 219–242.

Kraska, M. (2008). Assessment. In N. J. Salkind & K. Rasmussen (eds.), *Encyclopedia of educational psychology*. 2nd edition. Thousand Oaks, CA: SAGE, 60–65.

Leuenberger, T. (2001). Developing and using diagnostic and summative assessments to determine students' conceptual understanding in a junior high school earth science classroom. In D. P. Shepardson (ed.), *Assessment in science: A guide to professional development and classroom practice*. Dordrecht: Springer, 199.

Lindblom-Ylänne S. (2003). *Oppimisen psykologia ja ylioppilastutkinto*. [The psychology of learning and matriculation examination]. In A. Lahtinen & L. Houtsonen (eds.), Oppi osaamiseksi–tieto tulokseksi: Ylioppilastutkinnon 150-juhlavuotisseminaari. Helsinki, 38.

Lunetta, V.N., Hofstein, A. & Clough, M.P. (2007). Learning and teaching in the school science laboratory: An analysis of research, theory, and practice. In S. K. Abell & N. G. Lederman (eds.), *Handbook of research on science education*. New Jersey: Lawrence Erlbaum Associates, 393–442.

Martinez, M.E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34(4), 207–218.

McMahon, M., Simmons, P. & Sommers R. (2006). *Assessment in science: Practical experiences and education research*, Arlington, Va.: NSTA Press, 29.

McMillan, J.H. (2008). *Assessment essentials for standard-based education*. 2nd edition. Thousand Oaks, CA: Corwin Press, 6–38.

McTighe, J. & Ferrara, S. (1998). *Assessing learning in the classroom. Student assessment series*. Washington D.C.: National Education Association, 11–20.

Miller, M.D., Linn, R.L. & Gronlund, N.E. (2008). *Measurement and assessment in teaching*. 10th edition. New Jersey: Pearson Education, Inc., 1–287.

Murphy, P. & McCormick, R. (2006). Problem solving in science and technology education. In J. Gilbert (ed.), *Science education: Major themes in education*, volume II. London: Routledge, 186–214.

Nitko, A.J. & Brookhart, S.M. (2007). *Educational assessment of students*. 5th edition. New Jersey: Pearson Education, Inc., 1–260.

Osterlind, S.J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. 2nd edition. Boston: Springer, 30.

Pelton, T. & Pelton, L.F. (2006). Introducing a computer-adaptive testing system to a small school district. In S. L. Howell & M. Hricko (eds.), *Online assessment and measurement: Case studies from higher education, K-12, and corporate*. Hershey, PA: Idea Group Inc (IGI), 146.

Phelps, A.J. (1996). Teaching to enhance problem solving. *Journal of Chemical Education*, 73(4), 301–304.

Plake, B.S. (2005). Doesn't everybody know that 70% is passing?. In R. P. Phelps (ed.), *Defending standardized testing*. Mahwah, NJ: Routledge, 182.

Popham, W.J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: ASCD, 72–105.

Quellmalz, E. & Hoskyn, J. (1997). Classroom assessment of reasoning strategies. In G.D. Phye (ed.), *Handbook of classroom assessment*. San Diego, CA: Academic Press, 111–112.

Reid, N. & Yang, M. (2002). *The solving of problems in chemistry: The more open-ended problems*. Research in Science & Technological Education, 20(1), 83–98.

Rodriquez, M.C. (2002). Choosing an item format. In G. Tindal & T. M. Haladyna (eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation*. Mahwah, N.J.: Lawrence Erlbaum Associates, 213–232.

Salmio, K. (2004). *Esimerkkejä peruskoulun valtakunnallisista arviointihankkeista kestävän kehityksen didaktiikan näkökulmasta*. [Examples of national basic education evaluation programmes from the perspective of the didactics of sustainable development]. Joensuu: Joensuun yliopistopaino, 29–30, 60–69, 164.

Scheerens, J., Glas, C.A.W. & Thomas, S.M. (2003). *Educational evaluation, assessment and monitoring: A systemic approach*. London: Routledge, 100–110.

Stiggins, R.J. & Erter, J.A. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Portland, Or.: Assessment Training Institute, 99–100.

Tamir, P. (2003). Assessment and evaluation in science education: Opportunities to learn and outcomes. In B. J. Fraser & K. G. Tobin (eds.), *International handbook of science education: Part two*. Dortrecht: Kluwer Academic Publishers, 761–785.

Tarendash, A.S. (2006). *Let's review chemistry: The physical setting*. 4th edition. Hauppauge, N.Y.: Barron's Educational Series, 491.

Temiz, B.K., Taşar, M. F. & Tan, M. (2006). *Development and validation of a multiple format test of science process skills*. International Education Journal, 7(7), 1007–1027.

Uusikylä, K. & Atjonen, P. (2005). *Didaktiikan perusteet*. [Fundamentals of didactics]. 3rd edition.. Helsinki: WSOY, 71–73, 191–208.

Wakeford, R. (2003). Principles of student assessment. In H. Fry, S. Ketteridge & S. Marshall (eds.), *A handbook for teaching & learning in higher education: Enhancing academic practice*. 2nd edition. London: Routledge, 42–61.

Welch, C. (2006). Item and prompt development in performance testing. In S. M. Downing & T. M. Haladyna (eds.), *Handbook of test development*. London: Routledge, 304–305.

Woolfolk, A. (2007). *Educational psychology*. 10th edition. Boston, Mass: Allyn and Bacon, 560.