Multiple-Choice Questions for Teaching Quantitative Instrumental Element Analysis: A Follow-Up

Originally published in:
Journal of Chemical Education 100(10), https://doi.org/10.1021/acs.jchemed.3c00061

Supporting Information

Multiple-Choice Questions for Teaching Quantitative Instrumental Element Analysis: A Follow-Up

Gunnar Schwarz

Laboratory of Inorganic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland.

* corresponding author: schwarz@inorg.chem.ethz.ch

PART B

Content

1. Multiple-choice question types

The term multiple-choice question encompasses many types, which are generally distinguished based on their format. They usually consist of an introduction or stem, which can start with a description of the context or a scenario, but also contains the actual question or task. This is followed by a series of options from which students must select none, one, or all, depending on the question type. The options to be selected are called keys, and the others are called distractors. Table S1 provides an overview. For more information, interested readers are referred to Billings et al.[1], Haladyna[2], and Krebs[3].

Table S1. A selection of common multiple-choice question types with examples of close content proximity, taken and adapted from ref. [4]. Keys are marked.

Type and description	Example
A Known as a single-choice question or conventional multiple-choice question. Only one option is the key.	*Which is an absolute method?* ☐ *ICP-MS* ☒ *Gravimetry* ☐ *Flame-AAS* ☐ *ICP-OES*
B Multiple-choice question. All, some, one or none of the options can be keys.	*Which is an absolute method?* ☐ *ICP-MS* ☒ *Gravimetry* ☐ *Flame-AAS* ☒ *Titration*
Two-sided question Also known as "alternative choice (AC) question". True-false questions are a variety. Two contrasting options are presented of which one is the key.	Alternative choice question. *What kind of method is gravimetry*? ☒ *Absolute method* ☐ *Relative Method* True-false question. *Assess the statement "Gravimetry is an absolute method."* ☒ *True* ☐ *False*
K Complex questions. Primarily, four options are presented to be evaluated (1 to 4 on the left). This is done with secondary options, which contain selected permutations of the statements that are chosen instead of the primary options.	*Which is an absolute method?* *(1) ICP-MS* *(2) Gravimetry* *(3) Flame-AAS* *(4) Titration* ☐ *1 and 2* ☒ *2 and 4* ☐ *3 and 4* ☐ *only 2* ☐ 1 to 4
K' or kprime A variant of the B type questions and multiple true-false questions. Scoring considers all four lines as a set. Full marks (e.g. one point) are only awarded if all four keys are selected. Some award half marks if three out of four are selected. **X** X type questions have the same format as as kprime type questions, but each line is evaluated and scored individually, i.e. each set may be worth up to four points.	*Which is an absolute method?* *True* *False* ☐ ☒ *ICP-MS* ☒ ☐ *Gravimetry* ☐ ☒ *Flame-AAS* ☒ ☐ *Titration*

2. Course description

The course has been previously described, including the utilization of a classroom response system during lectures.[5] In brief, Element Analysis is part of the undergraduate courses Analytical Chemistry I and II, held each year at ETH Zurich, covering basics in quantitative analysis, such as the analytical process, sampling and sample preparation, and calibration as well as inductively coupled plasma optical emission spectroscopy (ICP-OES), atomic absorption spectroscopy (AAS), X-ray fluorescence spectroscopy (XRF), and inductively coupled plasma mass spectrometry (ICP-MS). The course consists of one 45 min-lecture per week and non-mandatory weekly worksheets. In 2019 and 2020 the enrolment was about 180 students (mostly majoring in Chemistry and Chemistry Engineering, Interdisciplinary Sciences, and Material Sciences) for autumn semesters (Analytical Chemistry I) and 120 students for spring semesters (Analytical Chemistry II, with the same majors but not Material Sciences). As of the institution's regulations, only attending the final written exam is required; there are no other graded or mandatory activities.

Most students take the combined written exam of both courses (2 h in total, 30 min for element analysis), which includes several multiple-choice questions, which in turn account for one quarter to one third of the total score of the quantitative element analysis part. The second exam part is a quantitative element analysis problem and a number of specific questions, including calculations (e.g. concentrations), prompts to provide appropriate sampling, sample preparation, choice of method/measurement, calibration, and typically an evaluation of an exemplary result.

Table S2. Details to Figure 1. Scope question exam results taken from the summer exam periods for the combined exam of the courses Analytical Chemistry I and Analytical Chemistry II. While other exams also included scope questions, the combined summer exam seem the most comparable. [a]For 2021, only three items were counted for grading. Item 3 was dismissed as ambiguous.

Year	Candidates	Item 1		Item 2		Item 3		Item 4		All key selections	All key selections, %
		Text	Key selections	Text	Key selections	Text	Key selections	Text	Key selections		
2018	86	How high is the tolerable daily intake of lead?	83 (97%)	How do silver nanoparticles affect bacteria?	85 (99%)	What is the emission of heavy metals from waste incineration?	85 (99%)	How high is the tar content in cigarette smoke?	24 (28%)	277	81
2019	111	Determination of lead in white wall paint	110 (99%)	Quantification of silver in nanoparticles	94 (85%)	Quantification of amides in zwieback	100 (90%)	Determination of the oxygen content in blood	85 (77%)	389	88
2020	76	How does the intake of 1 mg of arsenic affect the metabolism of mice?	71 (93%)	Determination of trace elements in a piece of jade jewelry	75 (99%)	Differentiation between calcium oxide and calcium phosphate	39 (51%)	What is the oxalate content in spinach?	71 (93%)	256	84
2021	101	Determination of the alcohol content in wine	98 (97%)	Determination of the gold content in copper ore	101 (100%)	Determination of the carbon content in coal	not applicable[a]	Measurement of blood sugar	99 (98%)	298[a]	98[a]
2022	79	Influence of the sulfate content on the taste of water	78 (99%)	How do TiO_2 nanoparticles affect aquatic organisms?	78 (99%)	Quantification of thallium in hair in suspected poisoning	79 (100%)	How much silver is in silverfish?	77 (97%)	312	99

3. Worksheet description

The worksheets were part of the course described above. Lectures were complemented with weekly worksheets, which were made available to the students as a pdf via the learning management system (LMS, Moodle) and comprises between three and five multiple-choice and open-ended questions (occasionally with follow-up questions). In later worksheets, essay style questions were also included, for which the students were prompted to develop a suitable analytical strategy for a given element-analytical problem. The number of questions usually allowed students to complete each worksheet within 20-30 min. This was monitored by asking students specifically for the approximate time required to complete the worksheet.

Students were invited to submit their solutions anonymously with their names to the lecturer/lecture assistant by the following week either on paper or via email. Corrections with feedback, i.e. allocations of points and further comments as required, were provided in the following week. Example sample solutions with explanations for each question, a brief comment on the overall performance of students (e.g. average points gained, time required) and occasionally specific comments to common problems encountered and clarifications were also provided. There was no grade bonus for students submitting exercises. Students were also strongly encouraged to contact the lecturer/lecture assistant personally to inquire clarifications on the topics, exercise questions, corrections etc. whenever required.

4. Details to: Generating distractors from short-answer questions

Rodriguez[6] reviewed studies on the equivalence of multiple-choice and free-response items and concluded a high correlation between the two modes in cases for which the question stems are equivalent. The workload of teaching staff often does not permit them to sufficiently analyze responses to OEQs for further use,[7] thereby limiting the exercise aspect of questions. Phrasing useful distractors is one of the major hurdles in MCQ development[2,4,8] and implausible distractors are the most common flaw of these questions.[9] Not only is it necessary, as pointed out by Gierl *et al.*,[8] for an item to be written by content experts,[8] but also requires insights into students' approaches to the question, e.g. common errors and potential misconceptions. This combination makes the development of MCQs a highly skilled activity, even more than for OEQs. On the other hand, the expedient model for MCQ development with student interviews as it is conducted for educational studies,[10] seems to expansive for most instructors.[4] As distractors are one of the main hurdles, a regular advice for instructors for multiple-choice questions (MCQs) development is to use responses from student to earlier open-ended questions (OEQs) to come up with distractors. It was also a continuous, but anecdotal, experience from correcting countless worksheets and exams over the years, that for many short-answer questions (opposite to essay questions) the collective of student responses fall into some category and may be used as options for MCQs.

During the correction of written worksheets in the Fall 2019 semester, students' responses to the short-answer OEQs were categorized on the side and the number of reply themes noted, such as recurring statements and phrases (Figure 3). At first, this was intended to gather semi-quantitative information on students' overall performance to particular questions and potentially to reconsider, rephrase, or abolish questions for a more effective use of students' and the lecture assistant's time. As it turned out, the information was useful to convert some of the short-answer questions into MCQs and

phrase distractors from incorrect replies. Whether or not to convert an OEQ into a MCQ, the following criteria were used:

- the question did not require students to provide a long text as a response,
- there was a sufficient range of replies to formulate at least two plausible distractors or a set of plausible justifications for two-sided (true/false) questions,
- it was not essential that students phrase a response themselves (as it is the case for some questions to practice for the exam), and
- on the same worksheet there are still questions which required short or essay answers.

To verify that the distractor design was indeed appropriate, the responses of these two modes were compared with the rates of acceptable responses.

Due to the COVID-19 pandemic, most of the lecture for the Fall 2020 semester had been prerecorded in the lecture hall and made available to students via the LMS during normal lecture hours and for later viewing.[11] The weekly worksheets were made available via the LMS "test" activity and completed online. Although this did not allow anonymous submission of the solutions, the lecture assistant took care to process and correct them as anonymously as possible.
Selected questions were put to half of the students (a randomly assigned group, an LMS function) as MCQs and as OEQs to the other half. The two groups with altering MC and OEQs were formed to reduce the time required for the students to respond and thereby retain or increase participation under the differing remote teaching circumstance and the time required for the teaching assistant/lecturer to correct the replies to OEQs, while retaining the flow of "feedback" from the students by appreciating those OEQs replies. Out of 13 worksheets during the semester, six were identical for both groups and seven included questions of the different types for the groups.

General Worksheet Observations. Compared to Fall 2019 the number of students submitting solutions to the exercises about in Fall 2020 doubled on average from 21 to 43 per worksheet, while the average points achieved was about 10% lower. The latter may be due to the fact that more students did not respond to all questions. The participation increase may be associated with the special remote teaching environment in 2020. There was only a marginal difference in the average self-reported time required to complete the exercises (from 26 min in 2019 to 27 min in 2020 per question set), while some question sets showed particularly large ranges, e.g. 10 min to 120 min for the same set of questions. Though the assignment to the two groups was carried out randomly, the number of students submitting the exercises of group A was about half of those in group B. Furthermore, it needs to be mentioned that at this university there is a time gap of several weeks between the lecture period and the exams. This is often referred to as the "learning phase" by students, i.e. exam preparation time, and used to revise not only the lecture material but also (as the case may be first) engagement with the worksheet questions to check their learning progress.

Examples of Modified Questions. Box S1 display examples of modified questions. Their context is explained and the outcomes from the different modes discussed below. All questions and options for MCQs are shown in *italic*. *N* denotes the total number of responses and an asterisk (*) the accepted

option for MCQs. Responses listed as "other", may contain additional or incoherent replies. However, depending on the specific response or argument, "other" responses may also be deemed acceptable.

The first questions demonstrate the process of modifying the question and phrasing distractors according to previous responses to an OEQ. "Sample preparation" was the topic of the third lecture and included (acid) digestions. The "rolling" of soil and similar samples in water or diluted acids was mentioned under chemical procedures along dissolving samples and digestions, but not explained in detail. In order to explore this further, the subsequent worksheet in 2019 contained the following question: *Explain briefly and in contrast to a sample digestion why a soil sample is "rolled" (continuous movement of a slurry) in water or diluted acid for several hours. How will the sample subsequently be processed, i.e. which part will be used for further analysis?*

There were 35 submitted replies by the students for this worksheet. The following list contains the summarized responses to this question with the number of mentions.

- Separate different layers/densities 2
- Homogenize sample 5
- Other 5
- Subsequent digestion of sample 7

  --
- Subsequently, use solution 20
- Subsequently, use solids 6

In this particular year, no student provided a response which referred specifically to “extraction” of the heavy metals from the sample. Furthermore, a substantial portion of the students did incorrectly mention that a digestion would be performed after "rolling" of the soil and/or that the solid would be used further for analysis.

For 2020, this question was rephrased and separated into multiple tiers, whereas two tiers were put to the students as either a MCQ or an OEQ and the remaining tier only as a MCQ:

The modified phrasing elicited more correct responses, whereas the rate of correct responses was similar for both question modes. As these are MCQs, this set could readily be used during the lecture via a classroom response system.

Box S1. Example questions. Keys or acceptable responses are labelled with an asterisk (*). The numbers of responses, *N*, is provided.

1. *For the determination of (e.g.) heavy metals in soil samples, the samples are "rolled" in water or diluted acid (i.e. the slurry is continuously moved over several hours).*

a) *What is the purpose of "rolling" the sample?*

Group B (MCQ, *N* = 39)

A*	*Extraction of dissolvable heavy metals.*	59%
B	*Homogenizing the sample.*	38%
C	*Separation of different particle densities and sizes.*	3%

Group A (OEQ, categorized responses, *N* = 24)

- Extraction of/dissolving heavy metals* 62%
- To mix, homogenize and/or crush [the sample] 29%
- Dilution 4%
- Dissolve organic substances 4%

b) *Which part is used for analysis after "rolling"?* (only MCQ)

		Group A (*N* = 26)	Group B (*N* = 39)
A	*Solids*	23%	15%
B*	*Solution*	77%	85%

c) *Will a digestion be carried out afterwards?*

		Group A (MCQ, *N* = 26)	Group B (OEQ, categorized responses *N* = 39)
A	*Yes*	58%	44%
B*	*No*	42%	46%
			(Other 3%)

References

1. Billings, M. A.; DeRuchie, K.; Haist, S. A.; Hussie, K.; Merrell, J.; Paniagua, M. A.; Swygert, K. A.; Tyson, J. *Constructing Written Test Questions for the Basic and Clinical Sciences*. National Board of Medical Examiners: 2016
2. Haladyna, T. M., Mc Formats. In *Developing and Validating Multiple-Choice Test Items*, 3rd ed.; Haladyna, T. M., Ed. Routledge: 2004; pp 67-96.
3. Krebs, R., Verschiedene Itemtypen: Wozu, Wie? In *Prüfen Mit Multiple Choice : Kompetent Planen, Entwickeln, Durchführen Und Auswerten*, 1st ed.; Krebs, R., Ed. Hogrefe: Bern, 2019.
4. Schwarz, G. Questions for Classroom Response Systems and Teaching Instrumental Element Analysis. *Chimia*. **2021,** *75*(1-2), 33-38; DOI: 10.2533/chimia.2021.33.
5. Schwarz, G. Interface Model and Implementation Framework for Classroom Response Systems. *J. Chem. Educ.* **2021,** *98*(6), 2122-2127; DOI: 10.1021/acs.jchemed.0c01339.
6. Rodriguez, M. C. Construct Equivalence of Multiple-Choice and Constructed-Response Items: A Random Effects Synthesis of Correlations. *J. Educ. Meas.* **2003,** *40*(2), 163-184; DOI: 10.1111/j.1745-3984.2003.tb01102.x.
7. Briggs, D. C.; Alonzo, A. C.; Schwab, C.; Wilson, M. Diagnostic Assessment with Ordered Multiple-Choice Items. *Educational Assessment*. **2006,** *11*(1), 33-63; DOI: 10.1207/s15326977ea1101_2.
8. Gierl, M. J.; Bulut, O.; Guo, Q.; Zhang, X. Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Rev. Educ. Res.* **2017,** *87*(6), 1082-1116; DOI: 10.3102/0034654317726529.
9. Breakall, J.; Randles, C.; Tasker, R. Development and Use of a Multiple-Choice Item Writing Flaws Evaluation Instrument in the Context of General Chemistry. *Chem. Educ. Res. Pract.* **2019,** *20*(2), 369-382; DOI: 10.1039/c8rp00262b.
10. Haladyna, T. M.; Downing, S. M.; Rodriguez, M. C. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Appl. Meas. Educ.* **2002,** *15*(3), 309-334; DOI: 10.1207/S15324818ame1503_5.
11. Schwarz, G.; Bleiner, D.; Gunther, D. On Video Lectures During Remote Teaching and Beyond. *Anal. Bioanal. Chem.* **2022,** *414*(11), 3301-3309; DOI: 10.1007/s00216-022-03983-y.