Stephan Bach (2020). Using multiple choice questions in STACK – reasons and examples. In: Contributions to the 3rd International STACK Conference 2020. TTK University of Applied Sciences: Tallinn, Estonia. DOI: 10.5281/zenodo.3945809

# Using multiple choice questions in STACK – reasons and examples.

Stephan Bach[1]

**Abstract:** Although STACK was developed to enable student-provided answers in computer aided assessment of mathematics, the system also offers a wide range of opportunities for different types of multiple choice questions (MCQ). The article points out reasons for both the use of MCQ in the assessment of STEM subjects and the use of STACK instead of other LMS question types to realize this. Scenarios of usage are illustrated by examples from mathematical bridging courses.

**Keywords:** STACK, multiple choice questions, input types, assessment, practising

## 1 Introduction

One of the main reasons for the development of the assessment system STACK (System for Teaching and Assessment Using a Computer algebra Kernel) has been that there are several problems coming along with the use of multiple choice questions (MCQ) in STEM subjects, such as the possibility of just checking the different options instead of actually performing the mathematical operation to be assessed ([Sa13]). STACK therefore enables the use of constructed response questions (CRQ) in computer aided assessment (CAA). Considering this it might seem surprising to have an article about the use of MCQ in STACK. But these questions also have their place in the assessment and practising of mathematics. And in fact STACK offers various features to implement them.

The examples in this article are taken from mathematical bridging courses for prospective engineering students. The courses were developed and proven within the joint project OTH mind, funded by the German Federal Ministry of Education and Research. They address non-traditional students who often have a vocational education and whose prior mathematical knowledge is far below the general expectations at the beginning of the studies. The participants of these courses are encouraged to complete regular online quizzes with STACK questions ([Ba19]). In the context of these quizzes two observations were made by the author:

- Sometimes students perform better on rather hard questions with constructed response (CR) than on supposedly easy ones with multiple choice (MC). So the latter seem to address different learning objectives, which are not necessarily a condition to answer

[1] OTH Amberg-Weiden, University of Applied Sciences, Joint project OTH mind, subproject #aufstieggestalten, Hetzenrichterweg 15, 92637 Weiden, Germany, s.bach@oth-aw.de

the CRQ correctly.

- Especially in the field of bridging courses there are students who consistently have a hard time using the correct input syntax. Some of these issues are difficult to avoid in the validation process. Even if one is using the input option "Check the type of the response" students might enter valid expressions of the computer algebra system (CAS) (e.g. Fig. 1). But in fact using the correct input syntax is only a secondary learning objective for these students. In a first step they ought to learn how to solve equations, how to factor terms or how to add expressions with fractions. And they are frustrated if they get score zero even though they managed the actual construct of a question.

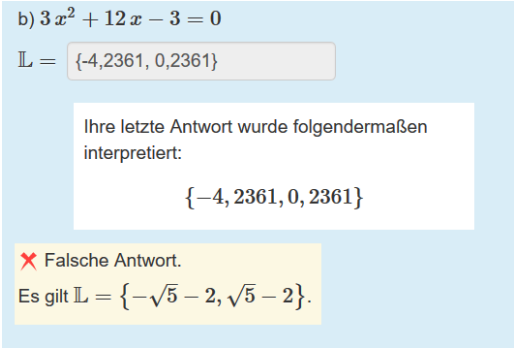


Fig. 1: Input issue – use of comma as decimal marker; screenshot in STACK 4.3

These two observations show that there are situations in which MCQ can play an important role in the learning process and that there are also problems which are related to CRQ.

Starting by summarizing the problems which go along with the use of MCQ in the assessment and practice of mathematics this article will point out reasons for both using these questions even so and for using STACK to implement them.

## 2 Problems with MCQ in the assessment of mathematics

There is a lot of research on the use of MCQ especially in high-stakes summative assessments[2]. This is mainly due to the higher economy of these questions. Many articles do not have a focus on CAA or STEM subjects though. Authors report on the following problematical aspects:

- **Reduced reliability due to guessing:** Multiple choice formats enable to find correct solutions by guessing or exclusion to some extent. Zimmerman and Williams point

[2] An overview is given by [LSK15].

out that this can reduce the reliability of multiple choice tests ([ZW03]).

- **Support of superficial learning strategies:** Students tend to use superficial learning strategies more often when preparing for a multiple choice based exam instead of a CRQ exam ([SC98]). Test-wiseness is discussed as an important influence on the performance in MCQ assessments ([TK 18]) and a possible source of construct irrelevant variance ([HD04]).
- **Assessment of reversible mathematical processes:** MCQ cannot be used for the assessment of reversible mathematical processes, such as differentiation/integration or expanding brackets/algebraic factoring ([SJ17]). Students will rather use the easier direct method to verify the options instead of actually performing the inverse process.
- **Assessment of creative mathematical tasks:** Classifications of mathematical questions contain several questions types which cannot be assessed by MCQ – many of them with a creative part, like finding a proof, giving an example or sketching a graph (also see [Sa13]).
- **Limited diagnostic validity:** Mathematical MCQ might have a reduced diagnostic validity due to the limited number of options. Kallweit mentions that the number of possible student answers can excel the number of distractors in a MC item by far ([Ka16]).
- **Gender bias:** There is a gender bias connected to the MC format regarding both the performance in ([BF 98]) and the attitude towards respective tests ([LR04]). Women tend to perform relatively lower in MC exams and to have less positive attitudes to this assessment type than men.
- **Great demand of experience and resources:** It needs experience, expertise and resources for quality management and review to develop high quality MC test items ([LSK15], [HD04]). Especially the reliability of small and locally developed tests might be reduced due to poor item quality.

Considering these problems it is a serious question, why one should use MCQ for the assessment of mathematics at all – especially in the context of CAA where MCQ do not have a significant economic advantage over CRQ.

## 3 Reasons for using MCQ in the assessment and practising of mathematics

Sangwin and Jones point out that “there are many contexts [in the assessment of mathematics] in which well designed MC items are more appropriate than other question formats” ([SJ17], p. 218). And also in settings of mathematical practising MCQ have their place.

### 3.1 Avoiding problems with the input syntax

Due to low computer skills, test anxiety or fatigue the input syntax can be an additional source of construct irrelevant variance ([HD04]). STACK is providing various features to support students who struggle in this area (e.g. validation, check of input type), but the correct input will still be an additional task, which might cause frustration especially for weaker students and beginners (see section 1). The occasional use of MCQ can give these students a lower hurdle to start with.

### 3.2 Assessing different educational objectives

While questions with CR in CAA often focus on mathematical skills like factoring out terms or computing a derivative, MCQ are suitable for different educational objectives (Fig. 2) as for example knowledge. The knowledge of key terms as well as procedural knowledge often is a condition to answer mathematical questions. Winter emphasizes that in the context of mathematical practising knowledge continuously needs to be reactivated, revolved and rearranged ([Wi84], p.9).

Which of the following functions are even?

a. $f(x) = \sin x$
b. $f(x) = \cos x$
c. $f(x) = \sin x + \cos x$
d. $f(x) = (\sin x) \cdot (\cos x)$
e. $f(x) = \sin(\cos x)$

Which of the following integrals can be solved by using integration by parts?

a. $\int \sin x \, \mathrm{d}x$
b. $\int \ln x \, \mathrm{d}x$
c. $\int x \ln x \, \mathrm{d}x$
d. $\int \sin(x^2) \, \mathrm{d}x$
e. $\int \sin^2 x \, \mathrm{d}x$

Fig. 2: Assessing conceptual and procedural knowledge; own illustration.

MCQ can also be used to assess conceptual understanding with high validity as done in concept inventories (Fig. 3).

Given an arbitrary function $f$, if $\lim_{x \to 2} f(x) = 5$ what is $f(2)$?

a. 2
b. 5
c. It must be close to 5.
d. $f(2)$ is not defined.
e. Not enough information is given.

Fig. 3: Assessing conceptual understanding; question from [PP20], own illustration.

### 3.3 Establishing high quality practising

Winter emphasizes the importance of variety in mathematical practising, which always needed new perspectives, new material and new connections ([Wi84], p.10). Using different question types can add to this variety. He also mentions the principle of problem orientation in practising, which means to give exercises a superordinate perspective[3]. Giving this perspective can sometimes be done by just adding a multiple choice question (Fig. 4).

Which of the following equations generally hold?

a. $a^2 + b^2 = (a + b)^2 - 2ab$
b. $(a - b)(b + a) = a^2 - b^2$
c. $(a + 3b)^2 = a^2 + 6ab + 3b^2$
d. $(x + y + z)^2 = x^2 + y^2 + z^2 + 2(xy + xz + yz)$
e. $(x - 2y)(x + 2y) = x^2 - 4y^2$
⋮

Fig. 4: Problem oriented practising; example from [Wi84], own illustration.

## 4 MCQ in STACK

### 4.1 Available STACK features

STACK is providing four input types which may be used to implement MC type questions:

- Checkbox
- Drop down list
- Radio
- True/False.

Except for True/False-questions the model answer is given as a nested list which is containing both the different options and the actual correct answer. The options can be mathematical objects like expressions or equations, plots, strings or even external images. The syntax may seem a bit difficult for beginners but there are additional helper functions provided. Details can be found on GitHub ([Mu20a]) and in the documentation ([Mu20b]).

### 4.2 Reasons for using STACK for MCQ

All the STACK input types mentioned above are available as easy to use, separate moodle

[3] This idea is also known under the terms of intelligent or productive practising ([BR08], [Le09]).

questions types. So why should question authors use STACK for MCQ?

### High quality MC formats

STACK is giving the opportunity to implement MC-formats which are recommended for university level assessment and which are suitable to reduce construct irrelevant variance due to guessing ([LSK15]) like multiple True/False-questions (Fig. 5)

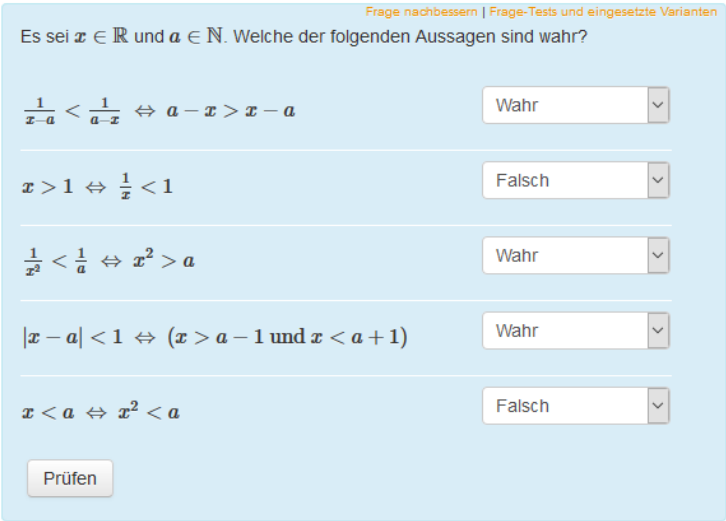


Fig. 5: Multiple True/False-question; screenshot in STACK 4.3.

### Effective implementation

Although it might be more difficult in the beginning it can be quite effective to implement MCQ in STACK – especially when using Maxima plots (Fig. 6). All the properties of the plots are stored in one single CAS variable, which makes it very easy to add changes or to adapt a question to another class of functions.

### Using typical STACK features

As with any other STACK input type one can use the features the system offers for randomization, differentiated feedback and grading. So it is possible to use randomized graphs (Fig. 6), to give specific feedback, also on options which are not chosen (Fig. 7), or to use specific grading methods for summative assessments.

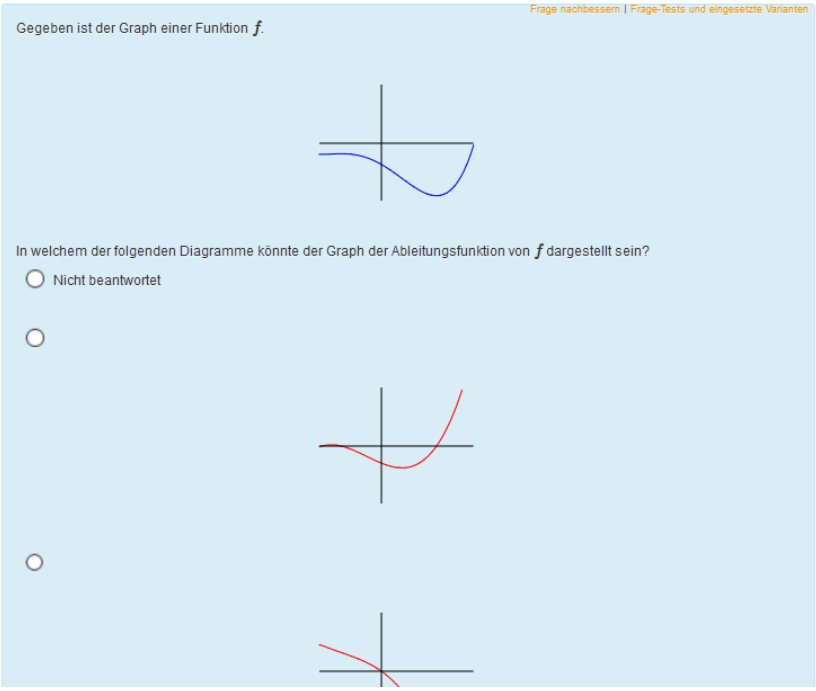


Fig. 6: Randomized maxima plots in MCQ; screenshot in STACK 4.3.

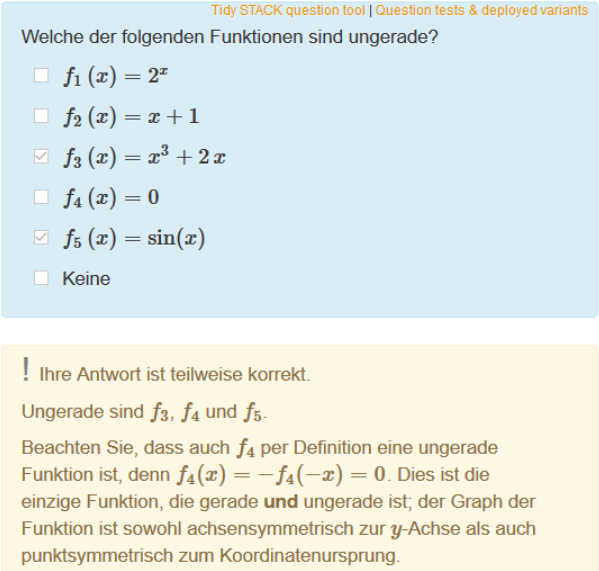


Fig. 7: Checkbox question with specific feedback; screenshot in STACK 4.3.

### Multi-part questions

In STACK one can combine MCQ with questions with algebraic or numerical input in multi-part questions. By doing so, it is possible to give an introduction or sensitization to a certain problem (Fig. 8) or to add the superordinate perspective which is needed for problem oriented practising (see section 3.3).

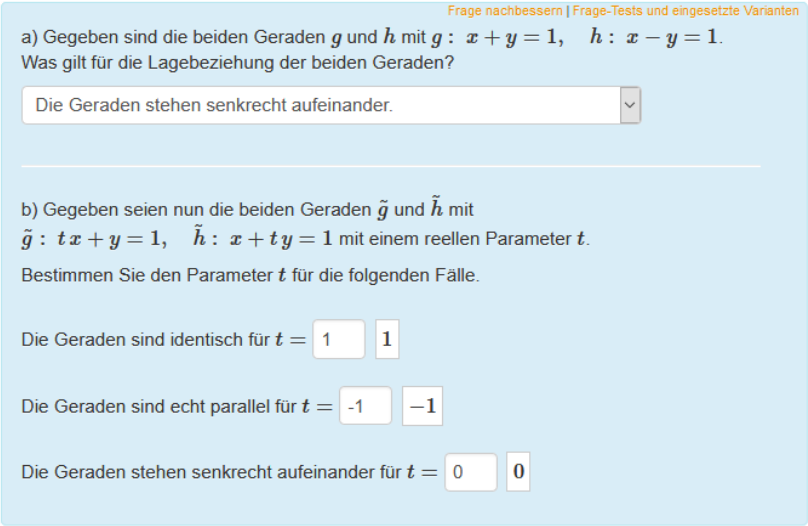


Fig. 8: Drop down list in a multipart question; screenshot in STACK 4.3

## 5 Summary and discussion

To establish high quality practice and assessment of mathematics it is helpful to use a mixture of different question types, including algebraic input, interactive graphs and also multiple choice questions. This will take into account that there are different types of learners, different educational objectives and different threats to reliability and diagnostic validity. Based on experiences the author made in the context of mathematical bridging courses especially weaker students could benefit from such a variety of input types. With its features for randomization, feedback, grading, plotting and multi-part questions STACK is a system which enables implementing MCQ of high didactic quality and combining them with other STACK input types.

Although many articles on MCQ, especially in summative assessments, have been published, there are several open questions. What is the role of MCQ in formative assessments and practising? Do the subject and the assessment method (paper based vs. computer aided) have an influence on the known differences between the question formats

(see section 2)?[4] And what is the relevance of possible input problems for the motivation and test performance of individual students, especially at the transition school-university? These questions should be object of further research.

## Bibliography


[Ba19] Bach, S.: STACK-Fragen zur Unterstützung der Selbstlernphasen in einem Studienvorbereitungsangebot für beruflich Qualifizierte. In: Contributions to the 1st International STACK conference 2018, DOI: 10.5281/zenodo.2563803, 2019.

[BF98] Birenbaum, M., Feldman, R. A.: Relationships between learning patterns and attitudes towards two assessment formats. In: Educational Resarch, 40:1, pp. 90-98, DOI: 10.1080/0013188980400109, 1998.

[Br08] Bruder, R.: Üben mit Konzept. In: Mathematik lehren 147, pp. 4-11, 2008.

[HD04] Haladyna, T. M., Downing, S. M.: Construct-Irrelevant Variance in High-Stakes Testing. In: Educational Measurement: Issues and Practice, 23 (1), pp. 17-27, 2004.

[Ka16] Kallweit, M.: Der Computer als Tutor – technikbasierte Diagnostik mit Freitextaufgaben. In: Institut für Mathematik und Informatik Heidelberg (Ed.), Beiträge zum Mathematikunterricht 2016. WTM, 2016.

[Le09] Leuders, T.: Intelligent üben und Mathematik erleben. https://home.ph-freiburg.de/leudersfr/mamo/Leuders_Intelligent_Ueben_2009.pdf, 2009.

[LR04] Livingston, S. A., Rupp, S. L.: Performance of Men and Women on Multiple-Choice and Constructed-Response Tests for Beginning Teachers. In: ETS Research Report, RR-04-48, 2004.

[LSK15] Lindner, M. A., Strobel, B., Köller , O.: Multiple Choice Prüfungen an Hochschulen? Ein Literaturüberblick und Plädoyer für mehr praxisorientierte Forschung. In: Zeitschrift für Pädagogische Psychologie, 29 (3-4), pp. 133-149, 2015.

[Mu20a] Multiple choice questions. https://github.com/maths/moodle-qtype_stack/blob/70da33bb7e7f832a5411abe0cfdeb43d7ece509d/doc/en/Authoring/Multiple_choice_questions.md, accessed: 03/06/2020.

[Mu20b] Multiple choice questions. https://stack2.maths.ed.ac.uk/demo2018/ question/type/ stack/doc/doc.php/Authoring/Multiple_choice_questions.md, accessed: 03/06/2020.

[PP20] PhysPort: Calculus Concept Inventory (CCI). Developed by Epstein, J., 2006. https://www.physport.org/assessments/assessment.cfm?I=57&A=CCI, accessed: 24/04/2020.

[Sa13] Sangwin, C.: Computer Aided Assessment of Mathematics. Oxford University Press, 2013.

[Sc98] Scouller, K.: The influence of assessment method on students' learning approaches:


[4] According to Livingston and Rupp the performance difference of women and men in CR tests depends on the subject and is lower in mathematics ([LR04]).

Multiple choice question examination versus assignment essay. In: Higher education, 35, pp. 453-472, 1998.

[SJ16] Sangwin, C. J., Jones, I: Asymmetry in student achievement on multiple-choice and constructed-response items in reversible mathematics processes. In: Educational Studies in Mathematics, 94, pp. 205–222, DOI: 10.1007/s10649-016-9725-4, 2016.

[TK18] Thoma,G.-B., Köller, O.: Test-wiseness: ein unterschätztes Konstrukt? Empirische Befunde zur Überprüfung und Erlernbarkeit von test-wiseness. In: Zeitschrift für Bildungsforschung, 8, pp. 63-80, DOI: 10.1007/s35834-018-0204-0, 2018.

[Wi84] Winter, H.: Begriff und Bedeutung des Übens im Mathematikunterricht. In: Mathematik lehren, 2, pp. 4-16, 1984.

[ZW03] Zimmermann, D. W., Williams, R. H.: A New Look at the Influence of Guessing on the Reliability of Multiple-Choice Tests. In: Applied Psychological Measurement, 27, pp. 357-371, 2003.