

# Computer-based pre-clinical assessment

Does the embedding of multiple-choice questions in a clinical context change performance?

## Computer-gestütztes Prüfen im 1. Studienabschnitt

Ändert die Einbettung von Multiple-Choice-Fragen in einen klinischen Zusammenhang die Prüfungsergebnisse?

• Martin R. Fischer<sup>1</sup> • Veronika Kopp<sup>1</sup>

*Zusammenfassung:*

*Nach der neuen Approbationsordnung muss das erste Staatsexamen folgende Charakteristika aufweisen:*

- eine interdisziplinäre Fächerintegration
- klinische Bezüge
- und Problemorientierung.

*In der vorliegenden Studie wurde eine Prüfung entwickelt, die diese Forderungen umgesetzt hat. Sie bestand aus Einfachauswahl-MC-Fragen, die sich auf fünf verschiedene, klinische Syndrome bezogen. In einer Version waren die Fragen zufällig angeordnet, in der anderen waren die Fragen dem jeweiligen Syndrom zugeordnet. Jeder Fragenblock wurde mit einer Fallvignette eingeleitet, die einen Patient mit den typischen Manifestationen des Syndroms beschrieb. Die Fragen konnten unabhängig voneinander beantwortet werden und enthielten keinerlei Hinweise auf die richtigen Antworten. Die Prüfung wurde mit dem computerbasierten Lernsystem CASUS durchgeführt.*

*Ein Ziel der Studie war es herauszufinden, inwiefern die Einbettung der Fragen in einen Kontext die Performanz der Studierenden beeinflusst. Es zeigte sich kein Unterschied zwischen den beiden Gruppen. Ein weiteres Ziel war die Beantwortung der Frage nach der Akzeptanz solcher Lern- und Prüfarrangements. Hinsichtlich des Lernens am Computer zeigte sich bei den Studierenden eine hohe Akzeptanz, während diese bezüglich computerbasierter Prüfungen zurückhaltender ausfiel.*

*Schlüsselwörter: fallbasiertes Prüfen, computerbasiertes Prüfen, Multiple Choice, vorklinischer Studienabschnitt*

*Abstract:*

*According to the new German Federal Regulations for medical education, the written preclinical national boards examination ("1. Staatsexamen") should focus on*

- a horizontal integration of all subjects
- a vertical link through clinical references and
- problem-oriented terms of reference.

*To fulfill these requirements, the question context has to be changed while the question format "single best answer" is mandatory and unchanged by the new regulations. Within the framework of this study an examination was developed, with respect to these guidelines. It consisted of one-best answer multiple-choice questions, which related to five clinical syndromes. In one version of the exam the questions were put together randomly without reference to the underlying syndromes. In the second version the exercises were ordered according to the syndromes. Every block of questions was coupled with a case description that included the typical manifestation of the syndrome through the presentation of a patient's case. The exercises could be solved independently of each other and they did not contain any clues to the correct solutions. The examination was carried out on the computer based system CASUS.*

*An important issue in this study was the question of whether the embedding of examination exercises in a clinical context has an effect on the results of the examination. It was shown that the set up of the examination had no influence on the examination results. Another question concerned students' acceptance of learning with and taking an examination on a computer. Regarding learning with a computer, the students' acceptance was high, whereas they were more reluctant to accept computerized examinations.*

*Keywords: case based assessment, computer-based assessment, multiple choice, undergraduate medical education*

## Introduction

It is a well-known problem in various content domains, that students have problems to apply their acquired knowledge beyond the context, in which it was initially learned. The acquired knowledge remains "inert" [2][14]. Problem-based learning is one of the instructional approaches, which takes account of this problem by assuming that transfer occurs more easily if the context, in which students acquire their knowledge resembles the problems

in which the knowledge should later be applied. Thus, problem-based learning has a growing impact in the existing medical curricula as it was done at the University of Munich [10]. The change of educational goals (learning should be more problem- than subject-oriented by the integration of relevant clinical content in an 'appropriate manner') led to a modification of the German Federal Regulations for medical education in 2002. It requires the interdisciplinary integration of clinical topics into the first two preclinical years of medical study [3]. This modification of learning objectives results in a change of assessment. In regard to assessment it is re-

<sup>1</sup> Klinikum der Universität München, Medizinische Klinik-Innenstadt, Schwerpunkt Medizindidaktik, München, Deutschland

quired since fall 2005 to "assess the scientific and theoretical basics in the context of clinical problem-solving with the focus on medically relevant content" (ÄAppO, 2002, Section 22, Paragraph 3 [3]). Therefore, the examinations had to be adapted to the new demands. In order to gain insights into a first implementation of these demands, an empirical study was carried out. For the German National Boards Part I examination ("1. Staatsexamen") the single-best-answer question format (selection out of five alternatives) is mandatory. However, is there an effect on performance, when a clinical context is provided?

The format of written national boards examinations in Germany is still pencil on paper. We chose a computer-based approach for two reasons,

- a) logistical concerns [4], and
- b) to study the acceptance of this format of assessment, as the German National Examination Institute (IMPP) is developing a computer-based examination system.

Computer-based assessment has been intensively tested for some years for its practicality and psychometrical quality. Furthermore, computer-supported examinations allow for new assessment strategies that integrate audiovisual media, use dependent questions and simulations as well as innovative types of answers [6][9][11]. In the USA and Canada federal summative examinations in an online format are now routinely implemented [5].

The purpose of this study had two parts. The first was related to the design of the MC-examination after the two pre-clinical years: a comparison between an exam version "with clinical context information" and a version "without clinical context information". Do these differences in design have an influence on the outcome performance of the examination? Secondly, we focused on the acceptance of learning with and taking an examination on a computer. Furthermore, we asked for the usability of the program as a prerequisite for accepting computer-based examinations.

## Methods

### • Sample group and design

As we wanted to answer the question whether providing a context will influence the outcome performance, we compared two groups: one without context information and one with context information. To answer the question concerning acceptance of online-examinations, the students answered a questionnaire after the exam.

A total of 40 medical students took part in the study. All participants were preparing for the Part I of the German National Boards examinations, as laid down by the old system, which takes place at the end of pre-clinical training. Students were voluntarily recruited, responding to a poster advertisement and received €20 for expenses. The age of the participants varied between 21 and 25 years; the average age was around 22 years (arithmetical average = 21,9). The study took place in July 2003 with participants in Düsseldorf, Munich and Ulm.

The students were randomly assigned to individual groups, in which each test person was given an account in advance that corresponded to one of the two groups.

### • The sequence of events of the study

Following a short introduction to the system CASUS, the students had 75 minutes time to answer 50 one-best-answer MC questions. Subsequently they were asked to complete a questionnaire about the general acceptance of learning with a computer, doing computer-based examinations and the usability of the system. The study was carried out two weeks before the preliminary medical examination which all of the test participants had registered for.

### • The realisation of the condition "with context information" versus "without context information"

The starting point for setting up online-exams for the first stage of the medical examination, following the new legislation, was the presentation of five clinical syndromes (Parkinson's disease, asthma, pancreatitis, pyelonephritis, AV-block grade 3) by the use of patient cases (see Figure 1).

Mr. Zimmermann, 67 years old, initially complained about aching limbs and a depressed mood disorder. His symptoms progressed slowly and a typical clinical picture for Parkinson's syndrome developed with impaired arm movements while walking, rigidity, tremor, a toppling gait, a monotonous low voice, and few or reduced facial expressions. The clinical investigation showed the cogwheel rigidity sign related to increased muscle tone. Tremors occurred at rest with a frequency of 4-7/sec that levelled off with intentional movements and during sleep. The supination and pronation movements were depressed (bradykinesia).  
The following questions 1 – 10 are related to this case.  
Although the clinical motor symptoms spasticity and rigidity differ in many aspects, they have one thing in common. That is:

- (A) **Both go along with hypertonicity of the affected muscles recognisable as an increased resistance against passive muscle distension.**
- (B) Both go along with pathological reflexes of the Babinski-type
- (C) Both are based on primary function impairment of the basal ganglia
- (D) Both show increased phasic reflexes as the patellar tendon reflexes
- (E) Both are based on an impairment of the pyramidal track.

**Figure 1: Example of context information: Case presentation of Parkinson's disease and sample MC-question (correct answer in bold type).**

One-best-answer MC questions (A-Type) pertaining mainly to the fields of physiology, bio-chemistry and anatomy were chosen from a supply of examination questions which the IMPP had used in previous exams. Only questions that demonstrated a thematic connection to the syndromes were selected. They did not contain any clues to their solutions; neither within a case, nor globally and they could be answered independently from each other.

In order to be able to test the effects of the new examination format on performance results, a second examination version was set up consisting of the same questions but randomly ordered and without patient case information. This version follows the format of the previous (traditional) examination system. One group of students was tested with the first version ("with context information") and another group with the second version ("without context information").

### • Instruments

#### CASUS

This study was carried out with the online learning system CASUS. CASUS is a case-based learning system which is integrated into the medical curriculum of various medical schools including the universities of Munich and Düsseldorf. CASUS has often been used as a means of examination [8].

In CASUS the questions must be answered in the order they are presented. It is not possible to click forward in the exam without having answered the previous questions. One can, however, click backward to change a previous answer [6].

### Examination

The examination was made up of 50 MC questions, which had been used in previous national examinations. All of the questions had five possible answers, of which only one was correct. A correct answer scored as one point. The questions were the same for both groups. On average, the students were granted 90 seconds time to answer one problem as in the National Boards exam. Thereby the students had 75 minutes to work through the examination.

### Questionnaire

A questionnaire of twelve items was designed for this study, which the students filled out after having completed the examination. They were asked how they accept computer-based examinations, learning with computers, and the user friendliness (usability) of the system. The questions were answered on a 5-point Likert-scale, whereby 1 meant "not at all" and 5 meant "absolutely right".

#### • Analysis

The evaluation was generated using the SPSS statistics programme Version 12. Cronbach's Alpha was used to determine the internal consistency of the multiple-choice test and the reliability of the questionnaire's scales. Differences between the two groups were calculated by using a t-test for independent samples.

## Results

#### • Reliability

The reliability of the 50 item MC test was .84 (Cronbach's Alpha). The questionnaire's scales reached scores between .49 and .91.

#### • Item difficulty level and item-total correlations

The questions were developed in an elaborate manner with a rigorous review process for quality assurance, as the IMPP provided them. However, not all questions reached the recommended difficulty level (between .2 and .8) and/or positive item-total correlations [1], as shown in table 1.

**Table 1: Item difficulty level and item-total correlations for all 50 questions**

question	1	2	3	4	5	6	7	8	9	10	11	12	13
Item difficulty	.98	.63	.68	.30	.35	.48	.40	.57	.50	.50	.68	.25	.57
Item total correlation	.06	.29	.43	.01	.10	.12	.34	.15	.07	.42	.24	.27	.26

question	14	15	16	17	18	19	20	21	22	23	24	25
Item difficulty	.60	.98	.48	.70	.53	.63	.15	.70	.75	.40	.70	.45
Item total correlation	.57	.29	.46	-.05	.59	.32	.33	.45	.38	.34	.35	.25

question	26	27	28	29	30	31	32	33	34	35	36	37	38
Item difficulty	.70	.68	.45	.33	.23	.53	.78	.63	.73	.68	.53	.53	.75
Item total correlation	.40	.27	.59	.32	.23	.40	.18	.37	.03	.28	.22	-.08	.37

question	39	40	41	42	43	44	45	46	47	48	49	50
Item difficulty	.45	.65	.25	.57	.50	.30	.35	.80	.83	.18	.33	.23
Item total correlation	.28	.43	-.11	.65	.37	-.16	.43	.29	.30	.17	.24	.39

#### • Differences between the two groups concerning performance

On average the students achieved 26.9 out of a maximum of 50 possible points with a standard deviation of 7.69. The group with context information reached 27.0 points on average; the group without context 26.7. The difference between the two groups was not significant ( $p = 0.908$ ). There was no effect ( $d = .04$ ) (see table 2).

**Table 2: Overview of the performance of the groups with and without context information: average values (standard deviations in brackets).**

with context information (n = 19)	without context information (n = 21)	Total (n = 40)
27.0 (7.73)	26.7 (7.84)	26.9 (7.69)

When reanalysing the data after eliminating 13 items for low item total-correlation ( $< .2$ ) or inadequate difficulty level ( $< .2$  or  $> .8$ ), the overall performance for the group with context information was 21.0 ( $SD = 7.17$ ) and for the group without context information 19.6 ( $SD = 7.62$ ), respectively.

#### • Students' acceptance

There was a high acceptance for using a computer-programme for learning. The students were also generally satisfied regarding the acceptance of the usability of CASUS, that is to say how the study was carried out and the user interface. Students, however, were more reluctant when to accept examinations on the computer (see table 3 for details).

**Table 3: Overview of the averages of the scales.**

	N	$\alpha$	M (SD) <sup>a</sup>	Min	Max
<b>Acceptance of learning with a computer programme</b>	37	0.72	4.0 (0.74)	2	5
Computer learning programs should be used more often in the curriculum.	37	-	4.1 (0.86)	2	5
Computer learning programs are a useful way of learning.	37	-	3.9 (.84)	2	5
I can imagine to learn with computer learning programmes in the future.	37	-	3.9 (1.06)	2	5
<b>Acceptance of doing an examination on a computer</b>	37	0.91	3.3 (1.12)	1	5
The online format of the examination was appealing to me.	37	-	3.8 (1.15)	1	5
Faculty-wide examinations should be preferably done online in the future.	37	-	2.9 (1.22)	1	5
I can imagine to take online examinations.	37	-	3.5 (1.35)	1	5
High-stakes exams like the National Boards should be preferably done online in the future.	37	-	2.8 (1.3)	1	5
<b>Usability of the assessment system</b>	37	0.49	4.3 (0.55)	3	5
The introduction in the test procedure was sufficient.	37	-	4.7 (0.52)	3	5
I had no problems working through the test on the computer.	37	-	4.4 (0.98)	1	5
The text was legible.	37	-	4.6 (0.69)	2	5
The screen design of the programme was satisfactory.	37	-	4.0 (1.12)	2	5
The software ran smoothly without technical problems.	37	-	3.7 (1.32)	1	5

<sup>a</sup> 1 = not at all 2 = below average 3 = average 4 = above average 5 = absolutely right

## Discussion

The further development of Part I of the German National Boards examinations, as stipulated by the ÄAppO, recommends the assessment of subject-related knowledge that should be interdisciplinary connected by embedding the questions in a clinical context. In comparison to the previous format of the National Boards examinations, this study could not show, that the new format with thematic ordering of the questions regarding a clinical context significantly affected students' test performance [13]. This may be attributed to the fact, that the questions primarily aimed at descriptive knowledge and the context gave no clear hint for answering the questions correctly. As one had to answer the questions without context as well, there was no strong connection between context and questions. Referring to the semantic/cognitive dimension of the dimensional context model [7], both, questions and case information are of reduced context (contrary to an enriched context), meaning that the tasks can be construed as simple.

Another explanation seems possible: Due to the thematic ordering and the context information students (with context information) could activate the relevant scheme for the following questions and had therefore an advantage. However, as they needed time to read the context, time for answering the questions was lost. Overall, they perform as good as students without context. To verify this explanation, further studies are needed.

The students performance in the MC-test was unexpectedly low, especially as the students had to pass their National Boards Part I exam a few weeks after the study. One explanation could be the poor total-item correlation and the high difficulty of some items. Eliminating these problematic items, the students' performance does not change much. Considering this, it is likely, that the students did not take the exam seriously enough or were not adequately prepared, despite of the near high-stakes exam.

Concerning the results of the questionnaire, we suspect a positive selection bias as the participants voluntarily enrolled in the study. These students may have been more motivated for educational changes on the one side and particularly open to the idea of electronic assessment on the other.

Taking this into consideration, the level of acceptance among participants towards learning with computers was high. However, the students were more reluctant in their acceptance of computerized summative examinations. This scepticism was even bigger in the context of computerized National Boards examinations. Some of the students had concerns about the required technical reliability and data safety of the examination system. One could expect scepticism amongst students with every new format. However, computer literacy should not play a major role as 99% of our medical students are well acquainted to the CASUS learning system [12].

In summary, this study shows, that the set up of the pre-clinical MC-examination with regard to the thematic ordering of the questions and their embedding into a clinical context had any influence on the performance of the students. The practicality of running a web-based examination system for summative assessment has been confirmed although more testing is required. Other written examination formats should be taken into consideration to ensure interdisciplinary integration of pre-clinical subjects into a clinical context such as dependant pick-n multiple choice questions. This format of examination could only be carried out in a computerized way, thus justifying the extra effort for online examination infrastructure, data safety, and system reliability.

## Acknowledgements

We are indebted to Dietmar Neumann and Christian Götz for their support in planning and conducting the study, to Petra Vogel for question selection and creation of case vignettes, to Inga Hege, Martin Adler and Matthias Holzer for technical support and to Sibyl Hermann, Magnus Müller, Hubert Liebhardt and Franz Ruderich for organizational support.

This study was financially supported by the German Ministry for Education and Research (BMBF) within the CASEPORT-Project (FKZ 08 NM 111).

## Corresponding author:

• Dr. med. Martin R. Fischer, Klinikum der Universität München, Medizinische Klinik-Innenstadt, Schwerpunkt Medizindidaktik, Ziemssenstraße 1, 80336 München, Deutschland, Tel.: 089/5160-2159, Fax.: 089/5160-2366  
fischer.martin@med.uni-muenchen.de

## References:

- [1] Bortz J, Döring N. Forschungsmethoden und Evaluation. 3rd ed. Berlin: Springer-Verlag; 2002.
- [2] Bransford JD, Goldman SR, Vye NJ. Making a difference in people's abilities to think: Reflections on a decade of work and some hopes for the future. In R. J. Sternberg & L. Okagaki (eds.), Directors of development: Influences on the development of children's thinking. Hillsdale, NJ: Erlbaum. 1991:147-180.
- [3] Bundesrat. Approbationsordnung für Ärzte (ÄAppO) vom 27.06.2002. Bundesgesetzblatt. 2002;Teil 1(Nr. 44).

- [4] DeAngelis S. Equivalency of computer-based and paper-and-pencil testing. *J Allied Health*. 2000;29:161-164.
- [5] Dillon GF, Clyman SG, Clauser BE, Margolis MJ. The Introduction of Computer-based Case Simulations into the United States Medical Licensing Examination. *Acad Med*. 2002;77:94S-96S.
- [6] Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J. A modified key feature examination for undergraduate medical students: validation threats and opportunities. *Med Teach*. 2005;27:450-455.
- [7] Koens F, Mann KV, Custers EJFM, Ten Cate OTJ. Analysing the concept of context in medical education. *Med Teach*. 2005;39:1243-1249.
- [8] Kopp V, Herrmann S, Müller T, Vogel P, Liebhardt H, Fischer MR. Einsatz eines fallbasierten Computerprüfungsinstruments in der klinischen Lehre: Akzeptanz der Studierenden. *GMS Z Med Ausbild*. 2005;22(1):Doc11.
- [9] Ogershok PR, Moore RS, Ferrari ND, Miller LA. An Internet-based paediatric clerkship examination. *Med Teach*. 2003;35:381-384.
- [10] Putz R, Christ F, Mandl H, Bruckmoser S, Fischer MR, Peter K, Moore G. Das Münchner Modell des Medizinstudiums (München-Harvard Educational Alliance). *Med Ausbild*. 1999;16:30-37.
- [11] Schuwirth LW, van der Vleuten CP, Stoffers HE, Peperkamp AG. Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Med Educ*. 1996;30:50-55.
- [12] Simonsohn AB, Fischer MR. Evaluation eines fallbasierten computergestützten Lernsystems (CASUS) im klinischen Studienabschnitt. *Dtsch Med Wochenschr*. 2004;129:552-556.
- [13] Wender KF. Semantische Netzwerke als Bestandteil gedächtnispsychologischer Theorien. In: Mandl H, Spada H. (Hrsg.): *Wissenspsychologie*. Weinheim: PsychologieVerlagsUnion, 1988.
- [14] Whitehead AN. *The aims of education*. New York: Macmillan, 1929.