ASSSESSING COMPETENCIES IN ENGINEERING DESIGN EDUCATION WITH AUTOMATIC EVALUABLE EXAM TASKS

Frederike KOSSACK, Eike UTTICH and Beate BENDER
Ruhr-University Bochum, Germany

ABSTRACT
The possibility of automatic evaluation in online exams offers the advantage of automatic evaluation compared to paper-based exams with manual assessment. Nevertheless, teachers and students have major concerns about digital exams e.g. automatic evaluable question types are easier for students, because they can guess or recognise the answer without knowing it. To analyse these concerns for Engineering Design Education this paper investigates to what extent can be found differences in the results between digital and paper-based examination formats when assessing the same learning outcomes. For three courses, the analysed data contains one data set of a paper-based examination with mostly open questions and rather big complex tasks one data set with results from several small automatic evaluable tasks for the same learning outcome per course. Based on the analysis, the paper discusses provides recommendations for automatic evaluable exam tasks in Engineering Design Education.

Keywords: Engineering design education, digital exams, automatic evaluable tasks

1 INTRODUCTION
Digital (or online) exam usage is growing in higher education [1]. The possibility of automatic evaluation in online exams offers many advantages compared to paper-based exams with manual assessment [2]. Students are graded equal, transparent and fair and teachers have a reduced assessment expense [3]. However, the technical capabilities of the question types for automatic evaluation limit the design and implementation of tasks in online exams, which might lead to different exam outcomes compared to paper-based exams.
Engineering Design Education is a key subject for the development of technical products in degree programmes like Mechanical Engineering [4]. Typical results of exam tasks in Engineering Design Education are dimension parameters and safety assessments of machine parts, and sketches or technical drawings that represent concept or design results. These results differ from other subjects and go beyond pure calculations, as in mathematics. Students contextualise and evaluate the results they calculate, e.g., in dimensioning of machine elements. For this reason, the development of digital exam tasks is specific for engineering design education and differs from other subjects like mathematics. Several authors published approaches for new tools in Engineering Design Education that automatically evaluate technical drawings or CAD models, but none of them is ready-to-use and they require manual adjustments (compare [5–7]).
This paper applies ready-to-use question types for Engineering Design Education, e.g., numerical and closed question types like Multiple-Choice. Students tend to find it easier to answer closed questions, because they can guess or recognise the answer without knowing it. Consequently, teachers must pay special attention when developing exam tasks with closed questions to secure assessment quality.
Learning outcomes are central in the development of exam tasks [8]. Bloom proposes taxonomy levels from 1 (e.g. “remember) to 6 (e.g. “create”) to assess learning outcomes [9]. Literature considers closed question types suitable for assessing learning outcomes on lower taxonomy levels [10] but approaches that use Multiple-Choice question for higher taxonomy levels exist. This paper investigates to what extend there are differences in the exam results in Engineering Design Education, if the same general learning outcome is assessed in paper-based exams compared to automatic assessment in digital exams.

2 DIGITAL EXAM TASKS IN ENGINEERING DESIGN EDUCATION

To initiate the comparison between paper-based and digital exam results, we prepared exam tasks presented in this section. At Ruhr-University Bochum, the learning management system Moodle is available. This software is widely used in Engineering Design Education in Germany [11] and offers different typical task types for automatic assessment options in the activity quiz [12]. In close questions, learners choose from given answers, e.g., in a Multiple- and Single-Choice question or assign elements of different lists to each other e.g. in Matching tasks [10]. In open questions, learners enter an answer freely [13]. For automated assessments, the student answer is compared with teacher answers. Moodle offers several question types to input numbers. *Numerical* is for numbers with limited digits. *STACK* allows numbers or mathematical equations. However, with these technical options it is only possible to a limited extent to transform the existing tasks from paper-based exams into automatically evaluable digital exam tasks. To address similar learning outcomes with the tasks in paper-based and digital formats we used the approach to split general learning outcomes on high taxonomy levels into specific tasks on different taxonomy levels (compare [14]). Exemplary exam tasks of the different contents in Engineering Design Education for undergraduate students are detailed for both exam formats in the following sections.

2.1 Tasks for methodical product development

A typical topic is the methodical product development beginning with the investigation of requirements, the definition of functions, the selection of operating principles and embodiment design. Figure 1 shows an analysis task with the creation of a non-hierarchical embodied model of a described technical system.

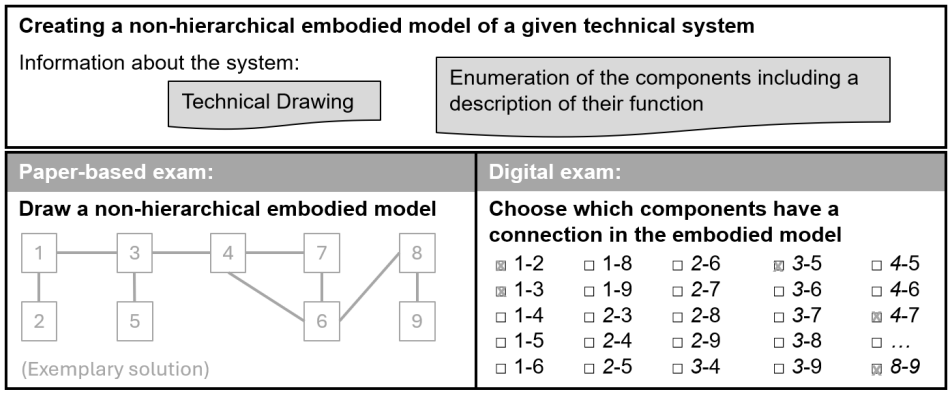


Figure 1. Simplified illustration of the comparison for an exam task about non-hierarchical embodied model

In the paper-based exam format students draw the model. In the online exam format students can draw the model on paper and then choose which components have a connection or they just pick the correct connections. Another exemplary exam task is about the ordering scheme shown in figure 2.

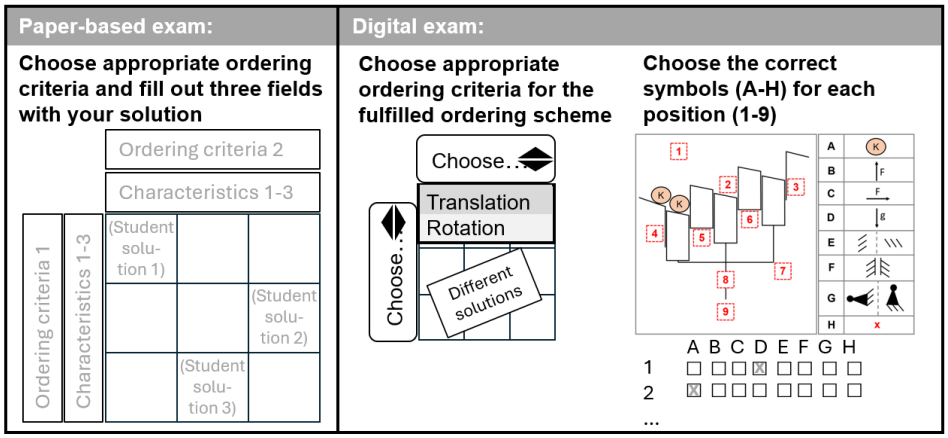


Figure 2. Simplified illustration of the comparison for an exam task about ordering schemes

In the paper-based exam students should define ordering criteria and create three fitting solutions. In the digital exams the associated learning outcomes are assessed with two tasks. One task to choose fitting ordering criteria for an already completed ordering scheme and one task to fulfil a given solution to assess, if students understand the technical system and can use notations properly. This example shows how the technical limitations for automatic assessable tasks lead to split large tasks in smaller ones.

2.2 Tasks for dimensioning machine elements and shaping of modules

The courses focus on dimensioning of individual machine elements e.g. bolts, bearings, shaft-hub-connections including the shaping of whole modules e.g. gearboxes. Typical paper-based exams consist of a use case with successive task parts (compare Figure 3). For assessing the same learning outcomes with existing question types in Moodle it is simpler to split the task in three independent tasks than to transform the paper-based task into a digital one without any changes. Figure 3 shows the transformation of one task about bearing calculation into three independent tasks. Two are implemented with closed questions (Multiple- or Single-Choice) and one with an open question. The open question with numerical inputs is implemented with STACK, which allows to evaluate anticipated mistakes with half points and subsequent errors with full points in the next steps. The detailed use of different types of questions especially STACK for calculating machine elements is explained in [15].

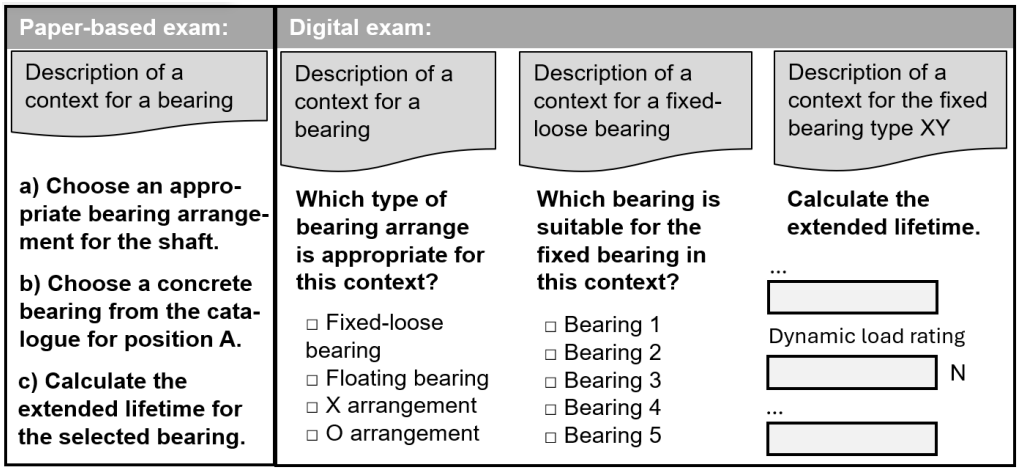


Figure 3. Simplified illustration of the transformation from paper-based exams and digital exams for calculating machine elements

3 PAPER-BASED AND DIGITAL EXAM RESULTS COMPARISON

3.1 Data collection and analysis

For investigating differences between paper-based and digital exam results, we compare summative exam result data sets from three different courses. One data set with exam results of each course is paper-based and the other one is digital. Conditions for the participants were the same in both data sets regarding permitted equipment, e.g., a non-programmable calculator, and a formula collection with needed equations for the calculation limited to a certain number of pages. All examinations were conducted at the university under supervision. The paper-based exams were held in a lecture hall and the online exams in computer rooms. Table 1 shows the concrete courses at Ruhr-University Bochum Germany and the year of data. All courses are part of the Bachelor's degree programme in Mechanical Engineering or Sales Engineering and Product Development. Machine Elements B (ME B) is part of the second study term, Machine Elements C (ME C) is part of the third, and Methodical Product Development is part of the fifth study term. The Machine Elements courses are mandatory for all students, Methodical Product Development only for students that specialise in Engineering Design. Due to the corona pandemic, we developed online exams which students took at home. However, they were subject to completely different examining conditions than paper-based exams. For example, students could use all kind of learning material like books or lecture sheets. When on-site exams were possible again, some courses got new digital exams with the same examining conditions as in paper-based exams and other courses had paper-based exams again. Therefore, the analysed data are from different years.

We tried to keep the time gap between the data sets of the various subjects as small as possible for the comparison.

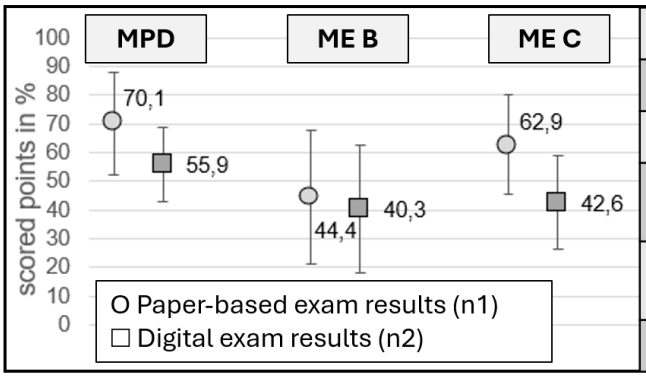
Table 1. Data sets used in the analysis with the number of data sets (n)

Course	Methodical Product Development (MPD)	Machine Elements B (ME B)	Machine Elements C (ME C)
Paper-based exam	Winter term 19/20 (n=130)	Summer term 23 (n=163)	Winter term 19/20 (n=342)
Digital exam	Winter term 21/22 (n=128)	Summer term 24 (n=130)	Winter term 21/22 (n=99)

We analysed the data sets with the statistic software tool SPSS version 29.02.0 [16]. Due to the varying maximum points and different numbers of tasks in the exams, we compared percentages achieved by the test groups. We defined results of the paper-based exam as test group 1 and results of the digital exam as test group 2. To identify differences between the two test groups, we compared mean and standard deviation. In addition, the data were investigated for statistically significant differences with the Mann-Whitney-U-Test. The test assumes that there is no statistic difference between two test groups that differ in terms of a characteristic [17]. In our case, the characteristic was the exam format (paper-based or online). The rejection range was defined by the significance level, which is usually set at 5 % [18].

3.2 Results

Figure 4 shows the analyses of the data. On the left is the initial comparison of paper-based exam results and digital exam results for the three courses. The mean of paper-based exams is higher than the mean of digital exams for all three courses. For MPD, the mean is 15 % higher, for ME B 4 % higher and for ME C 20% higher. The standard deviation for ME B is quite similar with 23,23 % and 22,17 %. It is slightly higher in the paper-based exam for ME C with 17,36 % compared to 16,18 % in the digital format. For MPD, it is about 5% higher in the paper-based exam (17,8%) than in the digital exam (12,8%).



	MPD	ME B	ME C
U	3231	9385	6401
z	-8,49	-1,67	-9,42
Median (n1)	74	47,06	66,12
Median (n2)	58,05	39,14	43,67
p	0,001	0,93	0,001

Figure 4. Initial comparison between exam results of three different courses with the mean and standard deviation (left) and results of the u-test for statistically significance (right)

A detailed investigation for statistically significance shows significant differences in the results between the test groups for MPD and for ME C. The significance p for MPD and ME C is smaller than 0,001. There is a statistically significance with a rejection range of 5%. There is no statistically significant difference between the two test groups of ME B with a significance value of p = 0,93. The right-hand side of Figure 3 shows the full results of the Mann-Whitney-U-test.

4 DISCUSSIONS

The results show that students scored better results in the paper-based exams in the courses included in our study. Based on our data, there is no assumption that digital exams with mostly closed question types generally lead to better exam results in engineering design education.

However, the size of the identified differences between paper-based exam results and digital exam results varies. To explain these differences, we looked at the design of the digital exams in detail. For all courses the digital exams assess the same learning outcomes as in the paper-based exams, but the design of the exams varies. Paper-based exams in MPD and ME C consist of large tasks with different parts, e.g., one example of a bearing and three different calculations are asked for in courses about machine elements or one technical system is described and a function model and creating solutions principles in a morphological box is required. The digital exams often have more small tasks, which are independent from each other. The similarity of digital exam task for ME B to the paper-based format of the course ME B is higher than of MPD and ME C. It consists of less independent small tasks than the digital formats for the other two courses. In MPD, the paper-based exam consists of three large tasks with three to seven different parts in each and the digital format consist of 34 independent questions and small tasks.

Analysing the completed results, students worked on nearly all subtasks in paper-based exams and the tasks in ME B in the digital format, whereas in the digital format for MPD and ME C, many students didn't give an answer even in close questions for all questions and tasks. Students didn't even try tasks, especially the ones at the end of the exam. Even if the paper-based exams and the digital exams assessed the same number of tasks and the same learning outcomes, we assume that students need more time to read and familiarise themselves with the use case of each respective question. Hence based on these data and detailed analysis of the completed exams, we suggest developing large tasks with one use case or example and design several tasks and questions about it.

However, these results are subject to some limitations. Firstly, in addition to the design of the exams varying framework conditions of the carried data sets need to be taken into consideration. Students in winter team 21/22 had a lot of their lectures in a digital format, while students in winter term 19/20 had all lectures on site. The regulations for registration of exams changed as well, participation in exams in winter term 21/22 were not counted as failed attempts for the limited numbers of attempts. So, we assume many students took the exam without preparing for it properly in this term. We do not have these differences in lecture format and examination regulations for our data sets in ME B. Secondly, the size of the test groups especially for ME C for the statistical analysis varies a lot. Thirdly, the year of studying influences students' scored results significantly (compare [19]) and the mean year of studying in the test group is not taken into consideration.

Further work to investigate the recommendation, that digital exam tasks should consist of several questions for one use case or rather large tasks in engineering design education, can include the analysis of more exam data sets from different courses and various universities. Maybe the same tendence can be seen in only paper-based exams as well, if one can compare results of exams with many small tasks with large tasks for the same course. In addition, a test design with the same test group of students performing both kind of exam types for the identical learning outcomes and comparing the results expands the insights. In addition, a questionnaire for this test group with items about the time management, challenges and difficulties in the exam formats would help to derive design rules for exams.

We assume that findings from other fields are only transferable to a limited extent due to the specific characteristics of Engineering Design Education with the understanding of a use context and the interpretation of the results in this specific context instead of pure calculations.

5 CONCLUSIONS

The paper shows the differences in exam results between paper-based exams and digital exams in engineering design education. The transformation of paper-based exam tasks into digital ones is limited by available existing technical question types for automatic assessment. These limitations lead to splitting complex tasks into several small independent ones.

The analysis of exam results of three courses about methodical product development and the calculation of machine elements shows that students score better in paper-based exams. So even digital exams formats consist of a lot of closed questions it is not easier for the students. The differences in the results vary in the three analysed courses and based on a detailed analysis of the digital exam design we assume that it is more challenging for students to work on many small tasks with different contexts than work on large tasks with the same context and different parts.

REFERENCES

[1] Heidkamp-Kergel B. and Kergel D. Diversitätssensible Didaktik mit digitalen Medien. MedienPädagogik 2022; 48:13–29. doi: 10.21240/mpaed/48/2022.06.04.X.

[2] Arnold P., Kilian L., Thillosen A. M. and Zimmer G. M. Handbuch. E-Learning: Lehren und Lernen mit digitalen Medien. 5th ed. Bielefeld: W. Bertelsmann Verlag; 2018. isbn: 9783825249656.

[3] Berkemeier M., Bilo A., Fischer Y., Fortmann M., Frommer A., Graf-Schlattmann M., Gollan H., Hahne N., Huth D., Kamin A.-M., Keller A. M., Kirberg S., Pahlke-Kullik K., Stegemerten B., Temps T. T., van Ackeren I., Wilde M. and Wrenger B. E-Assessment in der Hochschulpraxis. DuEPublico: Duisburg-Essen Publications Online, University of Duisburg-Essen, Germany; 2018.

[4] Albers A., Denkena B. and Matthiesen S. Einleitung. In Albers A, Denkena B, Matthiesen S (editors). Faszination Konstruktion: Berufsbild und Tätigkeitsfeld im Wandel; 2012. p. 13–4.

[5] Dillenhöfer F. Entwicklung einer KI-Objekterkennung für technische Zeichnungen im Maschinenbau. TU Dortmund; 2023. doi: 10.17877/de290r-24140.

[6] Fastabend J., Müller B., Roth D. and Kreimeyer M. Conceptualization of an artificial intelligence-assisted tutoring system for teaching technical drawing skills to undergraduate students. *Proc. Des. Soc*. 2024; 4:2835–44. doi: 10.1017/pds.2024.287.

[7] Hoppe L. V., Gembarski P. C. and Lachmayer R. Intelligent tutoring system as a tool of formative assessment in design education. In: *DS 110: Proceedings of the 23rd International Conference on Engineering and Product Design Education (EPDE 2021).* The Design Society; 2021. doi: 10.35199/EPDE.2021.40.

[8] Biggs J. B. and Tang C. S.-K. Teaching for quality learning at university: What the student does. 4th ed. Maidenhead: McGraw-Hill/Open University Press; 2011.

[9] Krathwohl D. R. A Revision of Bloom's Taxonomy: An Overview. *Theory into practice* 2002; 41(4):212–8.

[10] Mayer H. O. and Hertnagel J. Lernzielüberprüfung im eLearning. München: Oldenbourg; 2009. doi: 10.1524/9783486848984.

[11] Kossack F. and Bender B. Engineering design education at German universities: potential for a common basis to create personalized e-learning content. *Proc. Des. Soc*. 2024; 4:2885–94. doi:10.1017/pds.2024.292.

[12] Moodle Contributors. Test; Available from: https://docs.moodle.org/404/de/Test. [Accessed on 2024, 10 October].

[13] Middendorf W. Digitale Lernumgebungen - didaktische Möglichkeiten und praktische Fragen. null; 2022. doi:10.25656/01:24223

[14] Violante M. G., Moos S. and Vezzetti E. A methodology for supporting the design of a learning outcomes-based formative assessment: the engineering drawing case study. *European Journal of Engineering Education* 2020; 45(2):305–27. doi: 10.1080/03043797.2019.1622653

[15] Kossack F., Uttich E. and Bender B. Assessing competences in engineering design education with an automatically evaluable digital exam format. In *Proceedings of the ICED 2025*.

[16] IBM Corp. IBM SPSS Statictics. Armonk, NY; 2021.

[17] Rasch B. Quantitative Methoden: Einführung in die Statistik für Psychologie, Sozial- und Erziehungswissenschaften. 5th ed. Berlin: Springer; 2021. isbn: 978-3-662-63281-9.

[18] Döring N. and Bortz J. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. doi: 10.1007 /978-3-642-41089-5.

[19] Kossack F., Kattwinkel D. and Bender B. Praxisbericht: Potentiale adaptiven E-Learnings für die Konstruktionslehre. Journal of Technical Education (JOTED) 2024;12. doi:10.48513/joted. v12i2.281.