High Quality Assessment and Measurement of Competency in Engineering Education

Thorsten Braun (University of Stuttgart), Simone Loewe (University of Stuttgart), Nadine Wagener-Böck (TU Braunschweig), Freya Willicks (RWTH Aachen University)

Abstract—**Over the past years, more and more efforts have been made for the improvement of quality assessment in higher education. The introduction of constructive alignment in engineering education certainly brought significant changes to many institutions on the practical level. However, a quality measurement of the assessment outcomes continues to be a crucial aspect. Research based diagnostic modelling for the measurement of competencies has been introduced to university learning, applying their quality standards for psychological tests to student assessment. But it remains unclear, how such sophisticated models can be applied to situations where teachers are not primarily concerned with pedagogical research or social sciences. In addition, the role of online assessment and e-learning tools for testing and assessment is also constantly on the rise. This adds another dimension of possibilities but also challenges.**

This round table intends to explore some of the present practices, research results, and concerns with assessment in engineering education. What can be learned from this? Does it provide any helpful knowledge or tools for the everyday teaching and assessment? Or, maybe, are we already doing it anyway?

The primary goal is to inspire new perspectives on assessment and to facilitate a discussion of best practices. This round table is hosted by visiting higher education developers from German universities with focus on engineering and technology.

Index Terms— **assessment, competency, engineering education, e-assessment**

I. INTRODUCTION

The famous 'constructive alignment' triangle (figure 1) of a balanced teaching and learning design is well known in many countries and institutional settings [1, pp. 95-110]. According to this, the following steps have to be taken during the preparation of a course in order to 'constructively align' it: First, define the learning outcomes, then develop the assessment based on these outcomes and, third, select learning and teaching activities that enable students to achieve the outcomes [2]. The balancing of all three aspects offers a very flexible and powerful framework for higher education development.

However, it takes a lot of preparation time to design a course according to the principle of constructive alignment. For example, lecturers need to spend time to focus on the didactic orientation of their course. The benefits of these efforts are not always immediately apparent. Hence, to meet
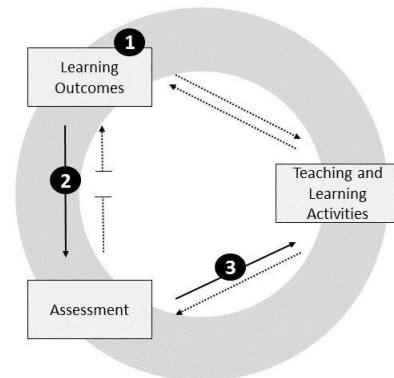
Fig. 1. Constructive Alignment Triangle, adapted from [1]

these challenges, standardized approaches have been suggested in the SoTL community [3]. Latest results in educational research on competency measurement and the rising possibilities of e-assessment offer the opportunity to further develop the work with learning outcomes in general and the design of examination questions in particular. The present paper aims at describing the possibilities resulting from these two topics, focusing mainly at the steps "learning outcomes" and "assessment" in the STEM field.

II. FIRST PERSPECTIVE: COMPETENCY MODELLING IN THE STEM FIELD

Competency modelling is the application of educational theory and research methods in order to analyze the structure of knowledge, abilities and skills in a certain group of people. To do so, elaborate methods have been developed in large scale school surveys like PISA [4]. Over the last years, first steps have been taken in Germany to apply competency modelling to higher education settings as well [5], especially in the crucial areas of the STEM fields with high numbers of students and dropout in the first year of study [6–8].

The underlying challenge here is that the results of students in examinations do not reflect the acquisition of competences that the course intended to impart. Usually, the expected learning outcomes define the structure of the test. Interestingly, research has shown that they are not always in line with the structure of competency a test actually measures. For example, your students may tackle several subject areas with a very similar strategy so that one competency is relevant for several different subjects. You might teach several topics with very different content, but the competency to solve problems in those topics is basically the same. Also, sometimes you might be surprised to learn that the 'easiest' tasks are significantly harder for some students than the 'challenging' ones. This information is

often blurred and veiled by simply summing up the points gained in a test to a single test score.

The intention of the approach we propose here is to fill the gap (figure 1) between the test results and the expected learning outcomes by doing a proper and robust analysis of the competency structure and the quality of the assessment design. We take test results as empirical data and apply these to more mathematical models in order to achieve a deeper understanding about what exactly the test did measure. Usually, such a modelling tries to give an answer to questions like the following:

a) What test items (assessment tasks) are easier than others and why? Are there clearly distinguishable levels of competency [9]?
b) For which competency level does the test provide a high degree of information [10, pp. 37]?
c) Do some test items measure something different for some groups of students [11, pp. 273]?
d) Does an assessment design fulfill scientific quality features [12]?
e) Are there any 'bad' test items with low discriminatory power [13]?

This inspired the idea to introduce competency measurement in the gap between test results and the learning outcomes in order to make the empirical structure of competency visible, to review the expected learning outcomes on a deeper level, and to learn more about how students are actually solving problems they encounter during a test. Some example results from competency modelling in the STEM field are:

a) students used essentially an unidimensional competency to solve problems from very diverse subject areas in thermodynamics, like ideal gas, heating and cooling machines, or circle process [14];
b) competency in general problem solving is crucial for the solving of mathematical problems [15];
c) prior knowledge is a very dominant factor for learning success in engineering mechanics [16];
d) mathematical competency has a major influence on all aspects of engineering mechanics, even pushing aside other aspects of competency [17].

Those results have to be considered as exploratory and only with limited, local validity; a unified, discipline specific model does not yet exist.

III. Methods for the Competency Modelling in the STEM field

The modelling of competency in the sense discussed mostly relies on probabilistic test theory and draws from log-linear models, especially from the group of the Item Response Theory derived from the so called Rasch Model [11]. Open source software packages are available in R [18]. Compared to classical test theory the estimated ability of a student is not based on weighted sum scores but takes into account how difficult any single task in the assessment actually is. This means, the modelling does not only estimate person ability but also item difficulty. The estimation has the advantage to place students and all test items on the same logit scale which makes both directly comparable. If the model fits the data, the universal metric scale allows for some analysis which is not possible in classical test theory; for example, the evaluation of personfit and itemfit statistics or the analysis of multidimensionality or differential item functioning.

IV. Second perspective: Developments in e-assessment

For some years now, e-assessments have been gaining increasing relevance. At many German universities they are already an integral part of everyday assessment routine [19]. Reasons for their use are the increased number of exams required by the Bachelor's and Master's systems as well as the increasing number of students. On the one hand, there are many benefits: The opportunities offered by e-exams are above all a simplified administration and reuse of questions through the possibility of using question pools. In addition, the evaluation of the exam is also carried out electronically, which saves a substantial amount of time. On the other hand, the time needed to create the exam questions for the first time is a challenge. In addition, the respective IT infrastructure needs to be in place, i.e. there must be an e-assessment system in use and there need to be enough PC workplaces for the examinees.

If you want to base your courses on the principle of constructive alignment, e-assessments offer a good opportunity since they allow you to approach the creation of exam questions in a structured way. How this works can be explained using the example of the German e-assessment system Dynexite: The system encourages the user to first formulate learning outcomes for his course by providing a standardized formulation guide. Three different levels of learning outcomes are distinguished according to Anderson and Krathwohl [2]. Each level symbolizes a certain complexity. In this way, the questions created are assigned to the previously formulated learning outcomes right from the start. Figure 2 gives an impression of the Dynexite system. It shows a part of the lecturer's view. After the system has automatically evaluated the exam, the lecturer can see at a single glance how many percent of the students passed (green bar) and failed (red bar) and how many percent didn't show up (grey bar). For more information about Dynexite visit https://www.dev.dynexite.rwth-aachen.de/ (in German).
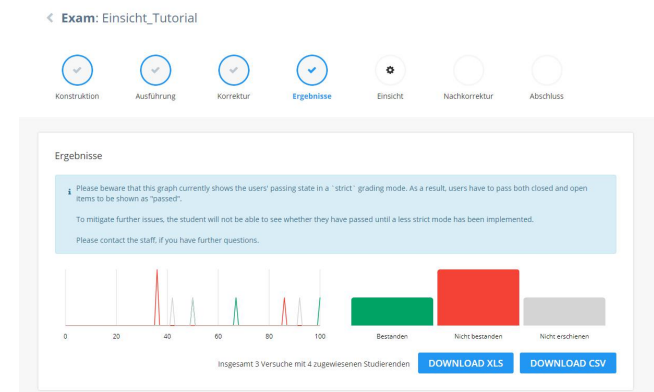

Fig. 2. Lecturer view in the Dynexite system, own illustration

V. Resulting opportunities and future vision

Those results inspire new visions for future assessment designs in engineering education, especially when combined with the rising possibilities of digitization. In theory, if someone would carefully design assessment tasks (test items) over consecutive years by using an e-assessment system like the mentioned Dynexite, a large pool of questions can be generated, sorted by complexity in the learning outcome. If, in addition, competency measurement is used to check for test quality and model requirements, the quality of the evaluation tasks could be greatly improved over time [20, pp. 156]. Such a pool could be used to tailor tests for specific levels of ability, subject areas, or even dynamic Computerized Adaptive Testing [21, 22]. Ideally, if the item pool and tests are well calibrated, even large groups of students could be provided with automated but qualitative feedback about their performance and weaknesses. Furthermore student placement tests or self-assessments could benefit from such calibrated item pools and tailored test designs.

So, this outlined future vision and perceived opportunity relies basically on the combination of competency measurement and the development of largely automated e-assessment online tools.

VI. Conclusion and questions

As innovative and efficient as the didactic work in such scenarios may be, in Germany this area of university research is still in its beginnings. At the moment, even the basic application is complex and always only a local solution. A continuous cooperation between educational researchers and subject teachers seems to be necessary. The round table envisioned with this paper would like to make its contribution to this field of research by discussing the following major questions:

a) What challenges arise in the outlined scenarios?
b) To what extent can parts of the outlined scenarios be transferred to the SoTL setting?
c) How can STEM departments of educational institutions provide meaningful support?

Acknowledgment

References

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

[2] L. W. Anderson, D. R. Krathwohl, and B. S. Bloom, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives : complete edition. New York: Longman, 2001.

[3] C. Jørgensen, A. Goksøyr, K. L. Hjelle, and H. Linge, "Exams as learning arena: A criterion-based system for justified marking, student feedback, and enhanced constructive alignment," in Transforming Patterns Through the Scholarship of Teaching and Learning: The 2nd European Conference For The Scholarship of Teaching and Learning, K. Mårtensson, R. Andersson, T. Roxå, and L. Tempte, Eds., Lund: Lund University, 2017, pp. 136–140.

[4] C. Sälzer and N. Roczen, "Assessing global competence in PISA 2018: Challenges and approaches to capturing a complex construct," International Journal of Development Education and Global Learning, vol. 10, no. 1, pp. 5–20, 2018.

[5] S. Blömeke and O. Zlatkin-Troitschanskaia, Eds., The German funding initiative "Modeling and Measuring Competencies in Higher Education": 23 research projects on engineering, economics and social sciences, education and generic skills of higher education students. Berlin, Mainz: Humboldt-Universität; Johannes Gutenberg-Universität, 2013.

[6] L. Gräfe et al., "Written University Exams based on Item Response Theory: MoKoMasch," in KoKoHs Working Paper, vol. 6, Current International State and Future Perspectives on Competence Assessment in Higher Education: eport from the KoKoHs Affiliated Group Meeting at the AERA Conference on April 4, 2014 in Philadelphia (USA), C. Kuhn, M. Toepper, and O. Zlatkin-Troitschanskaia, Eds., Berlin, Mainz: Humboldt-Universität; Johannes Gutenberg-Universität, 2014, pp. 30–34.

[7] O. Zlatkin-Troitschanskaia et al., Modeling and Measuring Competencies in Higher Education: Approaches to Challenges in Higher Education Policy and Practice. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.

[8] O. Zlatkin-Troitschanskaia, C. Kuhn, and M. Toepper, "Modelling and assessing higher education learning outcomes in Germany," in Higher Education Research and Policy, vol. 6, Higher Education Learning Outcomes Assessment: International Perspectives, H. Coates, Ed., Frankfurt: Peter Lang, 2014, pp. 213–235.

[9] E. Dammann, S. Behrendt, F. Ştef nic , and R. Nickolaus, "Kompetenzniveaus in der ingenieurwissenschaftlichen akademischen Grundbildung: Analysen im Fach Technische Mechanik," ZfE, no. 2, pp. 351–374, 2016.

[10] H. Wainer, E. T. Bradlow, and X. Wang, Testlet response theory and its applications. Cambridge: Cambridge University Press, 2007.

[11] W. J. Boone, J. R. Staver, and M. S. Yale, Rasch analysis in the human sciences. Dordrecht: Springer, 2014.

[12] T. Braun, "Die Klausur als Orakel?: Arbeitsergebnisse einer Klausurmodellierung in der Technischen Thermodynamik," zlw working paper, vol. 5, no. 1, 2018.

[13] D. Thissen and L. Steinberg, "Data Analysis Using Item Response Theory," Psychological Bulletin, vol. 104, no. 3, pp. 385–395, 1988.

[14] T. Braun, "Klausurmodellierung in der Technischen Thermodynamik mittels Item Response Theory: Ergebnisse einer hochschuldidaktischen Begleitforschung," Basel, Feb. 17 2018.

[15] I. Neumann et al., "Measuring mathematical competences of engineering students at the beginning of their studies," Peabody Journal of Education, vol. 94, no. 4, pp. 465–476, 2015.

[16] S. Behrendt, E. Dammann, F. Ştef nic , and R. Nickolaus, "Die prädikative Kraft ausgewählter Qualitätsmerkmale im ingenieurwissenschaftlichen Grundstudium," ZfU, no. 1, pp. 55–72, 2016.

[17] E. Dammann, S. Behrendt, F. Ştef nic , and R. Nickolaus, "Kompetenzniveaus in der ingenieurwissenschaftlichen akademischen Grundbildung: Analysen im Fach Technische Mechanik," ZfE, no. 2, pp. 351–374, 2016.

[18] A. Robitzsch, T. Kiefer, and Wu Margaret, Test Analysis Modules ('TAM'): Version 2.12-18. R Package. [Online] Available: https://cran.r-project.org/web/packages/TAM/TAM.pdf. Accessed on: Aug. 30 2018.

[19] M. Berkemeier et al., "E-Assessment in der Hochschulpraxis: Empfehlungen zur Verankerung von E-Assessments in NRW," 2018.

[20] F. B. Baker, The Basics of Item Response Theory, 2nd ed. United States of America: ERIC Clearinghouse on Assessment and Evaluation, 2001.

[21] H. Wainer, Computerized adaptive testing: A primer. London: Routledge, 2015.

[22] D. Magis, D. Yan, and A. A. von Davier, Computerized Adaptive and Multistage Testing with R: Using Packages catR and mstR. Cham: Springer, 2017.