

# Indicators for teaching assessment

Miriam Hansen<sup>1</sup>, Julia Beitner<sup>1</sup>, Holger Horz<sup>1</sup>, and Martin Schultze<sup>1</sup>  
<sup>1</sup>Goethe-University Frankfurt, Germany

This commentary on Schönbrodt et al. (2022) and Gärtner et al. (2022) aims at complementing the ideas regarding an implementation of DORA for the domain of teaching. As there is neither a comprehensive assessment system based on empirical data nor a competence model for teaching competencies available, yet, we describe some pragmatic ideas for indicators of good teaching and formulate desiderates for future research programs and validation.

*Keywords:* teaching quality, indicators, constructive alignment

## Differences research – teaching

Schönbrodt et al. (2022) and Gärtner et al. (2022) have written two target papers (TP1 and TP2, respectively) to advance an implementation of DORA regarding a more quality-based research assessment. We welcome these proposals and would like our commentary to complement them with thoughts on meaningful assessment of teaching. As both target papers, we focus on the situation of hiring and promotion in psychology.

To date, there is no comprehensive assessment system based on empirical data for the area of teaching or teaching competencies. It would be desirable to first develop a competence model for the area of university teaching, to adapt this in turn in a differentiated manner for different disciplines and contexts, and then to define criteria for the assessment of individual teaching competence with reference to this model. However, since hiring processes need to be improved rather sooner than later and this is an undertaking that can neither be realized in a short time nor by individual actors, we have focused in our following commentary on pragmatic indicators for the assessment of teaching.

These can be used as an intermediate step, as they are intended to help systematize the assessment of teaching in the short term, based on the information that is often already available or very easy to provide in application and hiring processes.

First, we describe how the problem of implementing valid measures of teaching quality differs from the facet of research. Thereafter, we propose pragmatic ideas for indicators of good teaching as first steps towards a more systematic and quality-based teaching assessment. Finally, some desiderates are specified for future research programs on higher education teaching quality and on the validation of our pragmatic indicators.

In contrast to the indicators available for the assessment of research quality which both target papers utilize, the assessment of teaching cannot rely on quality control to be carried out in prior steps to the same extent (e.g., by peer review). In the case of successfully acquired third-party funding and published manuscripts, the assessment of quality was performed by experts and aggregates of their appraisals are used as indicators in hiring and promotion. Such a system has not emerged for teaching. Using mere listings of courses taught does not contain any information about their quality, thus not providing insight into the academic valor of candidates. As such, it is merely an indicator of the characteristics of the positions previously held by a candidate. With regard to the lack of expert evaluation, even the results of student course evaluations, which are frequently listed as evidence, cannot be considered a substitute due to numerous possibilities of bias (Kornell & Hausman, 2016; Zabaleta, 2007). Therefore, the consideration of indicators for "activities in teaching" always requires an evaluation of their quality as well – analogous to the concrete exemplary attempt by Gärtner et al. (2022) in TP2. However, a uniform scheme for such an evaluation might be much more difficult to develop for the field of teaching, since local conditions and subject cultures have an important influence on the design of teaching even within psychology.

Currently, there are no established, quantifiable metrics or even indicators (cf. TP1; Wilsdon et al., 2015) for assessing output in teaching, such as the H-index. However, there are widely established practices for assessing teaching quality, e.g., in the context of hiring processes for permanent positions, that are equally questionable in terms of usefulness and quality: Usually, results of student teaching evaluations are referred to

as only evidence of teaching quality. Sometimes this is supplemented by a short teaching concept, occasionally also as part of a more comprehensive teaching portfolio, which then includes evidence of higher education didactic qualification or teaching awards. In principle, it is also possible and applied at some universities to ask for a teaching lecture or teaching samples as part of the job interview procedure for a permanent position (Meizlish & Kaplan, 2008). While these allow for quality assessment by peers and experts, they are often of limited standardization. Additionally, they are too broad in scope with regards to content while also being too narrow in the actual sample that is generated (e.g., a 20-minute sample lecture cannot appropriately depict competencies in activating teaching over the course of an entire semester).

While we believe that indicators for teaching quality currently used in hiring and promotion in psychology are not appropriate, there are enough possible data sources for assessing teaching performance and complementary ideas. Thus, we propose possible indicators below and describe our thoughts for an assessment of teaching competence based on these data sources and indicators.

### Description of possible indicators

#### Phase 1: Screening of previous work

In line with what is called “negative selection” in TP1 (p. 6, Figure 2), we propose an initial screening of previous work related to teaching. Of the possible data sources, published textbooks are perhaps most in line with the indicators proposed for the assessment of research qualifications. They undergo previous quality control and have readily quantifiable impact (e.g., citations, sales numbers). For the most part, this is also true for chapter contributions to textbooks. Further, published research on own teaching (following the so-called “scholarship of teaching and learning” approach, e.g., Hutchings et al., 2011) can also count as publications that inform about the merits in teaching. Additionally, as is done for research, third party funding for teaching projects should be used as an indicator of the willingness and ability to actively develop new approaches for teaching in higher academia. A third source of information, much less easy to quantify than the previous two, are open educational resources (OER), which have been identified as a key contributor to UNESCO’s Sustainable Development Goals (UNESCO, 2019) and are bound to be of increasing importance in modern approaches to teaching. These may range from material that is of textbook style and quality to more experimental and interactive materials such as ShinyApps or entire custom-

built websites. Because these often do not undergo an editorial process involving external feedback, the quality of these must be assessed individually during this phase. In terms of more “traditional” teaching materials, most courses are accompanied by online materials (often provided via a platform such as Moodle). These materials allow insight into structure, depth of content, use of diverse didactic methods, and the provision of additional in-depth information for a course taught in the past. This information may be supplemented by a written statement about the ideas underlying the design of the course.

A final area of previous work is additional didactic qualification, often represented by participation and certification of post-graduate academic development programs. Within the last decades, such programs have been widely implemented in many higher education institutions around the world (International Consortium for Educational Development (ICED), 2014) and such certifications are easy to assess and use as indicators. However, it should be kept in mind that the scope of qualification offered may vary depending on the location.

It is important to note that these suggestions we provide here should not be limited to traditional courses at universities, but should also include workshops and other forms of teaching. Workshops are often taught on a voluntary basis and aimed at teaching PhD students – therefore information about their quality may meaningfully supplement the picture generated by information about teaching undergraduates.

#### Phase 2: In-depth analysis of teaching quality

In line with “positive selection” of TP1 (p. 6, Figure 2), we also propose a second phase that inspects the teaching quality in addition to quantitative indicators.

To assess quality of teaching, we propose to first define different aspects of teaching that should be assessed. As a comprehensive model of higher education teaching competences is missing, we suggest as a starting point to use the subfacets of pedagogical/psychological knowledge (Baumert & Kunter, 2013). Since this model originated in the school context, it will certainly be necessary to make additions to the subfacets for higher education. For example, knowledge of conversation and supervision may be crucial for consulting of students, but also advising theses.

Further, we propose to make use of the concept of constructive alignment (Biggs, 1996) that postulates good teaching should start from the definition of intended learning outcomes which in turn should be aligned to the teaching and learning activities and to the assessment.

The candidates that were positively assessed in phase 1 can be invited to provide a detailed teaching concept for a previously taught course or a new one. Candidates should then elaborate on teaching objectives, didactic approach, content structure, examination form and questions. To complement the picture, candidates can present a teaching lecture or a teaching sample and – more important – an in-depth discussion and reflection of the teaching quality in defined subfacets. Situational interview questions can be used, for example "Using XY as an example, outline a) how you design teaching and exams and b) the extent to which you align teaching objectives, teaching/learning arrangements, and exams."

### Closing remarks

In this commentary, we outlined the difficulties in assessing teaching quality for hiring and promotion in psychology in relation to research quality. In addition, we propose a two-phase design which hiring committees could implement easily to improve the hiring process. Similar to TP1 and TP2, phase 1 focuses more on the quantitative aspects of previous work, while phase 2 allows an in-depth evaluation of teaching quality. Given that no model of higher education teaching competences exists yet, we propose several possible indicators, which can be used for assessing teaching quality in the meantime and which we consider to be an improvement in hiring and promotion compared to the status quo. However, we would like to stress that these indicators need to be empirically evaluated and assessed in terms of how well they actually reflect the quality of teaching. In addition, the predictive validity of these indicators about future teaching quality needs to be researched. We urge researchers in this field to take up the task of developing a scientific model of teaching competencies, so that the academic hiring processes can be further improved. Nevertheless, we consider this exchange and dialogue a promising sign for initiating actual change and an improvement in the assessment of candidates in psychology.

### Author Contact

Correspondence concerning this article should be addressed to Miriam Hansen, email: hansen@paed.psych.uni-frankfurt.de ORCID of authors: Miriam Hansen 0000-0003-1290-7630  
Julia Beitner 0000-0002-2539-7011  
Holger Horz 0000-0002-5173-0252  
Martin Schultze 0000-0003-1925-2403

### Conflict of Interest and Funding

The authors have no conflicts of interest or funding to disclose.

### Author Contributions

MH, HH and MS conceptualized the commentary. MH, JB and MS wrote the initial draft of the paper, and HH provided feedback on the first version. All authors contributed to the commentary and approved the submitted version.

### Open Science Practices

This article is a commentary and as such had no data or materials to share, and it was not pre-registered. The entire editorial process, including the open reviews, is published in the online supplement.

### References

- Baumert, J., & Kunter, M. (2013). The coactiv model of teachers' professional competence. *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*, 25–48.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022). *Responsible research assessment ii: A specific proposal for hiring and promotion in psychology*. <https://doi.org/10.31234/osf.io/5yexm>
- Hutchings, P., Huber, M. T., & Ciccone, A. (2011). *The scholarship of teaching and learning re-considered: Institutional integration and impact* (Vol. 21). John Wiley & Sons.
- International Consortium for Educational Development (ICED). (2014). *The preparation of university teachers internationally*. <http://icedonline.net/wp-content/uploads/2011/08/The-Preparation-of-University-Teachers-Internationally.pdf>
- Kornell, N., & Hausman, H. (2016). Do the best teachers get the best ratings? *Frontiers in psychology*, 7, 570. <https://doi.org/10.3389/fpsyg.2016.00570>
- Meizlish, D., & Kaplan, M. (2008). Valuing and evaluating teaching in academic hiring: A multidisciplinary, cross-institutional study. *The Journal of Higher Education*, 79(5), 489–512. <https://doi.org/10.1080/00221546.2008.11772114>

- Schönbrodt, F. D., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., Phan, L. V., Schmitt, M., Scheel, A. M., Schubert, A.-L., Steinberg, U., & Leising, D. (2022). *Responsible research assessment i: Implementing dora for hiring and promotion in psychology*. <https://doi.org/10.31234/osf.io/rgh5b>
- UNESCO. (2019). *Recommendation on open educational resources (oer)*. <https://unesdoc.unesco.org/ark:/48223/pf0000373755/PDF/373755eng.pdf.multi.page=3>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in higher education*, 12(1), 55–76. <https://doi.org/10.1080/13562510601102131>