

Book Review

Grading the College: A History of Evaluating Teaching and Learning by Scott M. Gelber

John Hopkins University Press, Baltimore, 2020.
248 pages. ISBN 9781421438160. \$44.95.

Evaluation has, in one way or another, always lied at the heart of the academic enterprise. And yet, in national and international contexts alike, evaluation is an ever hot topic. Who and what is being evaluated? For what purpose? To what and whose benefit? Who has the right to evaluate? What are the legitimate practices of evaluation? And, since we are at it, who gets to have a say in any of this?

Evaluative practices in higher education are nowadays typically imagined as an external imposition, performed for the purposes of, among others, transparency and accountability. Often, it is precisely the “outsider” character of such practices that provokes controversy, criticism, and frequently leads to them being interpreted as an attack on professional and institutional autonomy. Another frequently entertained belief is that the efforts to evaluate higher education emerged from market-inspired reforms, such as the new public management, which have characterised much policy making since the 1970s. Meanwhile, the earlier academic evaluation practices, however relevant they may have been for what came later, are—save for occasional mentions—routinely disregarded or, worse even, entirely forgotten.

In *Grading the College: A History of Evaluating Teaching and Learning*, Scott M. Gelber sets out to challenge some of the dominant beliefs about the origins and evolution of the practice in the United States. Gelber draws our attention to both deeper historical roots and the nature of faculty involvement in this history. Theory and practice of evaluation, he argues, are rooted in academic developments of the twentieth century and go all the way back to the interwar period and even earlier. Ever since, evaluation practices have been a subject of academic conversation, debate, and good deal of experimenting. However, social scientists, Gelber laments, typically begin their accounts with the 1980s—the decade when the discourse of “learning outcomes” took off, in the U.S. at least.

Gelber essentially asks: What happened before the 1980s and how did that matter? The book principally covers six decades, 1920-1980, and it documents continuous efforts to assess and measure the quality of undergraduate in-class teaching and learning, thus leaving out graduate education. It is organised in three parts and a concluding essay. The first part focuses on the history of teacher and teaching evaluation; the second on the evaluation of student learning; and the third part, also the shortest of the three, on the developments in both domains since the 1980. Once we are introduced to the rationale behind Gelber's ambition, Chapter 1 takes us a century back in time, where we find ourselves in the midst of heated debates about the "poor" state of teaching at U.S. higher education institutions during the 1920s. "Professors Are Poor Teachers," read an article headline. The discontent with the state of college instruction, as Gelber evidences throughout the first chapter, was widespread, effectively inspiring much discussion about how good teaching was to be defined, evaluated, and measured.

The debate on assessing teaching quality revolved around the efforts to substitute informal and often casual teacher evaluation practices with formalised and more systematic techniques, such as peer observation, student course evaluation, and measures based on teaching outcomes. Hearsay and gossip, as Gelber finds, played a decisive role in faculty promotion and tenure decisions. This was seen as a problem. We learn that, even as late as 1961, the American Council on Education acknowledged that rumours and casual conversations among faculty were still the most common sources of information about quality of teaching. As it becomes evident in the second chapter, most institutions ended up adopting student course evaluations as the default option. Student evaluations were relatively easy to produce, while not being perceived as a threat by the faculty: the reviews were mostly positive, and perhaps critically, they did not play an important role in faculty promotions. This, Gelber shows, changed towards the end of the 1970s, leading to the student evaluations being increasingly challenged by the faculty across the country.

While telling a history of teaching evaluation, which Gelber does carefully and engagingly, he also introduces us to some of the tensions which have historically defined and continue to define the discussions on the evaluation of teaching and learning. The tension between the inward-looking improvement and outward-facing accountability—as two somewhat divergent purposes of evaluation—emerges as a critical one. This tension becomes increasingly more explicit as the book progresses, although one is left with an impression that Gelber could have gone a step further and exploit its intricacies for the purposes of theorising the relationship between the university and society through the lens of evaluation.

Many true gems are contained in the first part, valuable not only as pieces of history, but also as historical mirrors to the present-day debates. One such gem is the variety of evaluation techniques considered across institutions, some of which originated outside higher education. Gelber tells us about the so-called “man-to-man instrument,” originally devised by the psychologist Walter Dill Scott for the U.S. military, which invited colleagues to rank one another. Corporate solutions, although much resisted by many among the faculty, were actively advocated by external parties such as the Carnegie Foundation which played anything but a minor role in the history of college evaluation. Students and their associations were increasingly more involved in these conversations over the years. By the 1950s, Gelber asserts, more than half of the college course evaluation programmes in the country were initiated by students. Gelber does a fine job of weaving students’ role in the history of evaluation into the broader historical progression from students as learners to students as consumers. Scholars interested in the changing role of students throughout history have much to gain from these insights.

The evaluation of student learning is the subject of the second part of the book. Gelber explores the history of standardised testing in some depth, while focusing on the support and resistance to the practice throughout decades. This is followed by a brief history of rubrics, surveys, and rankings, which were with various degree of success considered as alternatives to testing. This story of learning evaluation is, expectedly, also shaped by the aforementioned tension between inward-looking improvement and outward-looking accountability. However, compared to teaching, the history of learning evaluation seems to have been much more shaped by accountability discourses. This is likely a reason why various external constituencies with interests in higher education are given more attention in the second part of the book.

Although acknowledged from the beginning, the distinction between the evaluation of teaching and learning, on the one hand, and the evaluation of higher education institutions, on the other, remains rather elusive in the first part of the book. Only when the history of accreditation is unravelled towards the end of the second part does the author bring institutional evaluation to the mainstage. The reason for this could be that the history of accreditation is, in a way, a story of enduring professional and institutional resistance to agendas and pressures coming from various “outsiders.” Here Gelber neatly illustrates the effort colleges and universities put into retaining control over student testing, in particular when it came to the potential uses of the data thus collected. Higher education institutions were especially keen on ensuring that no public comparisons of institutions were produced, in particular if they were to be based on measurements.

As the capacities for large-scale assessments expanded, the pressure to focus on measurable student outcomes, but also on other measurable aspects of academic performance, was becoming more

intense. Here, historical advances in social sciences and computer technology played an important role in shaping these developments. Gelber draws our attention to, among others, psychometrics, which inspired thinking on student learning in the 1920s; to IBM developing machines which could automatically score multiple-choice exams in the 1930s; and to the growing prominence of human capital theory the 1950s and 1960s, which influenced much of policy thinking about economic aspects of higher education. It is tempting to think that a more systematic account of some of these broader historical developments could have helped the author make an even stronger case for an evolutionary approach in the study of evaluation. Nonetheless, the account Gelber provides is uniquely insightful.

One interesting aspect of Gelber's history of evaluation is that it is intertwined with the history of higher education studies in the United States. This is not explicitly addressed in the book, and probably rightly so as this would go beyond the book's purpose. However, it is worth noting that the evaluation practices Gelber talks about have roots also in academia's own longstanding interest in its own work: "Beginning in the 1920s and inspired by the growing influence and prestige of statistical analysis," Gelber writes, "educational 'measurement' emerged as a specialty that promised to yield reliable evaluations of teachers and students" (pp. 8-9). During the 1920s and 1930s, a growing number of universities established offices for institutional research. By the 1960s they became a standard feature of higher education institutions across the nation. These offices made use of social-scientific methods and increasingly more elaborate data-collection and analysis practices. In this sense, they epitomised a faith in the power of collecting, systematizing, and tabulating data for the purposes of improving quality and performance. Although he draws on a wealth of research produced from the 1960s onwards to document various findings regarding teaching and learning, Gelber misses the opportunity to explore the structural links between then already well-established institutional research tradition and the emergence of higher education studies as a distinct field of inquiry beyond particular institutions, which arguably also played a role in the history of evaluation.

Gelber seems acutely interested in disentangling the role of the faculty. In doing so, he meticulously examines contested terrains. He finds that faculty and administration were less likely to oppose the evaluation practices which were developed closer to their domain of influence, such as within "homegrown" examination programmes and regional accreditation bodies, than to those coming from non-academic third parties, most notably the Carnegie Foundation for the Advancement of Teaching and federal bodies. This may strike us as little surprising. Yet such criticism would miss an important point Gelber is making by drawing attention to these contested terrains—one which begs their consideration in a contemporary context. Higher education is always, anywhere, a site of an ongoing struggle, not only between universities and various third parties looking for ways to hold them

accountable, but also between academic and organisational dimensions transcending the “university-society” divide. Gelber paints a rather complex historical picture which shows how, over time, *organisation* gains in prominence, a trend much aligned with the accountability discourse. These developments made the way for what scholars would later recognise as the rise of organisational actorhood in higher education (Krücken & Meier, 2006).

Grading the College could be read as a history of two intertwined processes. On the one hand, it tells us about a growing importance of evaluation of teaching and learning, accompanied by more attention and resources sent in the direction of ever improving the methods. On the other hand, it is a story of academia’s loss of control over evaluation. In this sense, the book speaks to Espeland and Sauder’s *Engines of Anxiety* (2016), which focuses on rankings, but nevertheless speaks to this historical outcome. As Healy aptly put it in his review of Espeland and Sauder’s book: “At the heart of an academic ranking system is the experience of having one’s own knife turned back upon oneself, and finding that it still cuts like it used to” (2017, p. 519).

The “knife” metaphor resonates closely with Gelber’s narrative. He also seems a little irritated by the fact that academics shy away from acknowledging their historical role in this state of affairs. But, as we can see from the concluding essay, he is not dismissive; rather, Gelber takes off his historian hat to weigh in on the evaluation of teaching and learning, which he sees as one of “the greatest unsolved problems of academia” (p. 156). We are not surprised to learn that there is no magic formula and that each solution is a trade-off of sorts. For example, Gelber argues, an approach which seeks to cater for standardised quality control and external accountability is not likely to provide good guidance for faculty and administrators. In a reverse situation, the evaluation methods that the faculty would approve of and find most meaningful for professional development are not likely to interest external and often uninformed audiences, which tend to prefer simple yet flawed tools such as rankings. The methods which required a lot of faculty effort, regardless of their ethical and methodological validity, would neither enjoy broad popularity.

It is somewhat regrettable that Gelber decided not to include graduate departments. One can, of course, see many reasons how that would complicate matters, perhaps even take away something from the argument as presented. However, I cannot help but wonder how that would have shaped the story. For example, the most influential pre-1980s reputational rankings—the instrument Gelber sees as “emblematic of external evaluation” (2020, p. 106)—were precisely about graduate departments. Perhaps the best-known example is the so-called *Cartter Report*, published by the American Council on Education in 1966. It is, however, not as widely known that the idea to rank graduate departments came from the National Science Foundation, which had been considering hiring

a private research group for the task. The ACE vehemently opposed this, stressing that “*nobody* [emphasis in original] ought to play around with evaluating the graduate programs who was not themselves responsible to the institutions” (Allan M. Cartter quoted in Dolan, 1976, p. 26). Seen in this light, today much-resented reputational rankings were used by the ACE as a way to shield universities from external influence. As we would witness later on, however, the direction of the developments was such that it was probably only a matter of time when such shielding would eventually not be possible anymore, for undergraduate and graduate studies alike.

I have learned a great deal from this book. Gelber’s narrative is compelling, rich in detail, and some parts of the book are real page-turners. As a true historian, he is meticulous about the sources and does a true detective work in the archives. The book is an invaluable resource, not only for historians of evaluation in education, but for anyone grappling with the evaluation of teaching and learning, as well as with various aspects of institutional evaluation. Policy makers, university administrators, and not least teachers and learners have much to gain from reading this book. Gelber indeed writes about the United States and he does not seem much concerned with what was going on elsewhere. Yet it is naturally tempting to examine these developments against the history of evaluation practices in other parts of the world. Without any doubt, much could be learned from such an endeavour, especially considering the fact that the problems at the heart of this excellent book transcend national boundaries and speak to some of the most pressing issues in higher education and higher education research today.

Jelena Brankovic

Faculty of Sociology, Bielefeld University

Email: jelena.brankovic@uni-bielefeld.de

References

- Dolan, W. P. (1976). *The Ranking Game: The Power of the Academic Elite*. University of Nebraska Printing Service. <https://eric.ed.gov/?id=ED129131>
- Espeland, W. N., & Sauder, M. (2016). *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. Russell Sage Foundation.
- Gelber, S. M. (2020). *Grading the College: A History of Evaluating Teaching and Learning*. JHU Press.
- Healy, K. (2017). By the Numbers—Wendy Espeland and Michael Sauder, *Engines of Anxiety: Academic Rankings, Reputation, and Accountability* (New York, Russell Sage, 2016). *European Journal of Sociology / Archives Européennes de Sociologie*, 58(3), 512–519. <https://doi.org/10.1017/S0003975617000315>

Krücken, G., & Meier, F. (2006). Turning the University into an Organizational Actor. In G. S. Drori, J. W. Meyer, & H. Hwang (Eds.), *Globalization and Organization: World Society and Organizational Change* (pp. 241–257). Oxford University Press.