Review

Exploring high impact scholarship in research on student's evaluation of teaching (SET)

CrossMark

Pieter Spooren [a, *], Frederic Vandermoere [b], Raf Vanderstraeten [c], Koen Pepermans [a]

[a] Faculty of Social Sciences, University of Antwerp, Belgium
[b] Department of Sociology, University of Antwerp, Belgium
[c] Department of Sociology, Ghent University, Belgium

ARTICLE INFO

ABSTRACT

Student evaluation of teaching (SET) is the most common way of assessing teaching quality at universities. Since the introduction of SET procedures at the start of the previous century, thousands of research studies on the validity, reliability and utility of SET were written. By means of a citation analysis on journal articles included in Google Scholar, Scopus, and the Social Science Citation Index (Web of Science), this paper aims at mapping the high impact studies, the leading researchers, and the key journals in this research field. The results indicate that, although we find considerable overlap between the three data-bases, a number of high impact journal papers are not included in all three databases. Furthermore, the analysis reveals three main topics in the SET literature: the use of SET, validity issues concerning SET, and the construction and validation of SET instruments. Also, it is shown that many high impact studies were written by only a few researchers, with Herbert Marsh as the leading author. Although some of the most impactful studies date back to the 1960s, it's coming of age situates in the seventies. Since then SET became increasingly visible. The high proportion (25%) of impactful articles since 2000 indeed suggests a trend of continuous growth in SET research. At the same time, the historical knowledge in the form of classic studies on SET lives on through the many citations in recent studies.

© 2017 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author. Department of Social Sciences, University of Antwerp, Belgium.
  E-mail address: pieter.spooren@uantwerpen.be (P. Spooren).

Nowadays, student evaluation of teaching (SET) is used as a measure of teaching performance in almost every institution of higher education throughout the world. This widespread use is largely due to the apparent ease of collecting the data and presenting and interpreting the results (Penny, 2003). Whereas SET in the early days had a mainly formative character, in the 1970s it quickly became an important instrument in faculty personnel decisions as well (Galbraith, Merrill & Kline, 2012). More recently, SET-procedures are included as a key mechanism in internal quality assurance processes to prove an institution's performance in accounting and auditing practices (Johnson, 2000). The main purpose of SET is thus threefold: (a) improving teaching quality, (b) appraisal exercises (tenure/promotion decisions), and (c) institutional accountability (demonstrating adequate procedures for ensuring teaching quality) (Kember, Leung, & Kwan, 2002).

Since the introduction of the first 'teacher rating scale' in 1915 (Marsh, 1987; Wachtel, 1998), SET instruments and procedures have however provoked heated discussions among the various stakeholders. In general, their concerns include (a) the differences between the ways in which students and teachers perceive effective teaching, (b) the relationships between SET scores and factors that are unrelated to "good teaching" (Centra, 2003; Marsh, 2007), (c) SET procedures and practices (i.e. the contents of SET reports, the depersonalization of the individual relationship between teachers and their students due to the standardized questionnaires and respondents' anonymity, the competency of SET administrators, the low response rates, etc.), and (d) the psychometric properties of SET instruments. In the wake of the implementation of SET procedures, a new research theme has emerged in the field of educational psychology and educational sciences. The 'golden age of research' on SET however must be situated in the 1970's, when a lot of research dealt with issues concerning the utility and validity of SET (Centra, 1993).

This has probably much to do with what the American sociologist Harold L. Wilensky (1964) famously described as the professionalization of everyone. With the advance of experts, he argued, the profession became a pervasive occupational model that was spreading well beyond traditional professional areas, such as medicine and law (Wilensky, 1964). However, the distinction between professionals and their public was also redefined. The public became more influential. In this sense, the growing interest in SET signals an evolution towards more inclusive forms of (higher) education. The interest in professionalization is complemented by an 'upgrading' of the public. SET signals an active concern with inclusion, with the expansion of an upgraded membership of students in higher education. The forms of 'acceptance' of students in the relevant setting of higher education changed: students have become entitled to evaluate the professionalism of their teachers, they are given a 'voice' in decision-making processes. The relationship between professors and students can no longer remain an asymmetrical, hierarchical one. SET procedures are an indication of the 'upgraded' roles for the "laity", for the public (similar to, for example, the demands for internal democracy in the churches, for giving the laity a 'voice' in the ecclesiastical decision-making bodies).

At the moment, many comprehensive reviews on this subject are available (see a.o. Costin, Greenough, & Menges, 1971; Marsh, 1984, 1987; McKeachie, 1979; Onwuegbuzie, Daniel, & Collins, 2009; Spooren, Brockx, & Mortelmans, 2013; Wachtel, 1998). These studies often present a state of the art of the research on SET, thereby offering a structured overview of the (both classic and recent) research themes and study results in the field. But these reviews are often written from a particular point of view. The inclusion of a particular study in such a review mainly depends on its authors as they decide on several grounds (with respect to content and/or more technical aspects) whether or not they will mention a certain study in their work (i.e., research theme, methods, suited for the review theme, publication date, publication type, etc.). Still, this is only one way to evaluate the impact of a certain study on its research area. Another possibility, which is of a more quantitative nature, is performing a citation analysis to map out the most important studies in the field (Moed, 2005). The scholarly impact of a certain study is then measured by means of calculating the number of times this study article is cited by other studies. In this view, highly cited studies can be considered as studies that receive(d) much attention and, as a consequence, are in the center of the debate.

The quantification of scholarly impact has been significantly affected by the development of academic search databases such as Web of Science (WoS). Over the past decades WoS evolved from an early idea by Eugene Garfield to one of the most widely used citation-indexing services. Today it counts several thousand of institutional subscriptions and contains more than a billion citations (Clarivate Analytics, 2017). Around the turn of the century also other databases entered the market of bibliometric analysis. In 2003 Elsevier, one of the world's largest publishers of academic journals, launched another subscription-based tool named Scopus. One year later (November 2004) Google introduced Google Scholar (GS), a freely accessible software system that searches for scholarly literature on the Internet. A number of studies soon appeared on the relative pros and cons of each of these interdisciplinary, and other more specialized, citation indexes such as PubMed (in the