



On the credibility of the judge A cross-classified multilevel analysis on students' evaluation of teaching

Pieter Spooen*

Department of Social and Political Sciences, University of Antwerp, Sint-Jacobstraat 2, B-2000 Antwerp, Belgium

ARTICLE INFO

Article history:

Received 9 November 2009
Received in revised form 3 February 2011
Accepted 8 February 2011
Available online 1 March 2011

Keywords:

Students' evaluation of teaching (SET)
Teacher evaluation
Validity studies
Grading leniency
Multilevel analysis

ABSTRACT

This paper presents the results of an exploratory cross-classified multilevel analysis on students' evaluation of teaching (SET) in a medium-sized Flemish university. The revalidation of the evaluation questionnaire and the existence of a Global factor, the 'teacher professionalism' factor, which accounts for 52% of the variance in 7 (out of 12) SET-scales are discussed. Cross-classified multilevel analysis including student, course, and teacher characteristics shows that at the student level class attendance, students' age, course grade, and exam period are statistically significant predictors of SET (measured as the score on the teacher professionalism factor). At the course/teacher level, SET scores had a statistically significant association with rank of instructor. However, these indicators explain little variance in SET-scores (PRV is maximum 6.3%). The implications of this study for future SET-research are discussed.

© 2011 Elsevier Ltd. All rights reserved.

Setting the stage

Since the introduction of instructor rating scales at the start of the previous century (Marsh, 1987), students' evaluation of teaching (SET) has become a widespread phenomenon on college campuses for many years. Serving both formative (e.g., improvement of teaching) and summative (e.g., tenure and promotion decision-making) goals, SET is considered a key indicator in quality monitoring (Penny, 2003). However, the implementation of SET resulted in an explosion of research projects aimed at obtaining more insight in (a) the (im)possibility of constructing valid and reliable evaluation instruments (mostly paper-and-pencil questionnaires), (b) the validity and reliability of student perceptions of teaching, (c) the (ethical) use of student evaluations when mapping teachers' instructional skills and (d) the effects of SET on the improvement of education. Although there is consensus that student ratings are positively associated (correlations of .40 and more) with supervisor, colleague, and observer ratings (Beran & Violato, 2005; Marsh, 1987), most research deals with the question whether or not SET is 'biased' by student, course or teacher characteristics (for extensive overviews see e.g., Chonko, Tanner, & Davis, 2002; Marsh, 1987; Spencer & Schmelkin, 2002; Wachtel, 1998). 'Bias' then should be considered as the situation wherein "a student, teacher, or course characteristic affects the evaluations made, either positively or negatively, but is unrelated to any criteria of good teaching, such as increased student learning" (Centra & Gaubatz, 2000, p. 17). In this

case, evaluation results are misinterpreted or misused (Haladyna & Hess, 1994). Especially when SET is used for personnel decisions by college boards, educational administrators and policymakers, the possibility of bias still remains an extremely important research topic in the field of SET. It is certain that most experts and leading researchers are convinced of the validity of SET, since they find no or only little influence of possible biasing factors (Centra, 2003). Nevertheless, the previous research has not led to conclusive evidence with respect to the absence of threatening factors, since "one might suppose that the research studies on [student] ratings are similar to many other studies in education: conflicting, confusing and inconclusive" (Kulik, 2001, p. 10).

Next to the bias-question, several authors indicated the existence of halo-effects in SET – the general impression of a person influencing evaluations of that person on other conceptually similar or dissimilar attributes (Apodaca & Grad, 2005; Shevlin, Banyard, & Griffiths, 2000; Spooen & Mortelmans, 2006). Such 'halo-factors', explaining a lot of the variation in SET on various dimensions of teaching, might lead to incorrect interpretations of evaluation results. Shevlin et al. (2000) for example found a higher order factor, which they call the 'Charisma' factor, explaining 69% and 39% of the variation in their 'lecturer ability' and 'module attributes' factors, respectively. On the other hand, these findings partly show that, although student ratings are considered to be multidimensional (Marsh, 1987), students give similar ratings across a lot of evaluation items. In other words, it seems that SET have a multidimensional structure but are compatible with a very strong general underlying factor (Apodaca & Grad, 2005).

Although SET must be considered a multilevel phenomenon, only few studies used multilevel techniques when examining

* Tel.: +32 2655360; fax: +32 2755793.

E-mail address: pieter.spooen@ua.ac.be.

factors that might influence or bias a student's rating behaviour (Griffin, 2004; Nasser & Hagtvet, 2006; Pagani & Seghieri, 2002; Rampichini, Grilli, & Petrucci, 2004; Ting, 2000; Wendorf & Sheldon, 2005). SET indeed are the result of students' perceptions of the educational environment, but these perceptions can, as mentioned above, be coloured by both student and teacher/course characteristics. However, most studies only focus on one level and thereby ignore other levels. Or worse, higher level characteristics (mostly teacher or course characteristics) are brought down to the first level and thus considered characteristics on the student level. This leads to both statistical (spurious 'significant' effects) and conceptual problems (better known as *ecological fallacy*, i.e. analyzing data on one level and taking conclusions at another level) (Hox, 2002, p. 3). Second, the hierarchical structure of the population is denied: when using evaluations from several courses, some students are more the same than others (because, for instance they attended the same courses or are in the same educational program). Individual observations of teaching skills thus are not completely independent (Nasser & Hagtvet, 2006). This paper makes a contribution to the research literature on SET by executing an exploratory cross-classified multilevel analysis on SET-scores from a middle-sized Flemish university. Therefore, a number of known and previously unknown possible biasing characteristics on the student and the course/instructor level are used. The next section focuses on recent findings with respect to the variables used in the present study.

Recent findings and the present study

Student characteristics

Perhaps the most studied student characteristic in SET-research is the relation between *students' (expected) grade* and SET-scores. Since Feldman (1976, 1997) reported (as a result of an overview of the literature) the existence of a modest but statistically significant association (somewhere between .10 and .30) between (expected) grading and SET, there is an ongoing debate whether or not this relationship should be considered a proof for the validity of student perceptions of teaching since the early seventies (Griffin, 2004; Gump, 2007; Marsh & Roche, 2000). In the former case, it is suggested that good ratings reflect good learning (Marsh, 1987): students who learned a lot during the course (and thus probably received higher grades) reward their teacher for his/her educational activities. Better teachers make their students work harder and learn more, which leads to better learning outcomes. In the latter, some authors suggest that teachers can 'buy' better SET by giving higher grades (Greenwald & Gillmore, 1997; Krautmann & Sander, 1999; McPherson, 2006), which possibly has consequences in terms of grade inflation (Langbein, 2008). In this case we could speak of serious bias. A third hypothesis suggests the association between pre-existing student characteristics (such as subject interest or motivational aspects) and both grades and SET, which makes the grade 'effect' somewhat false (Marsh, 1987). In the present study, students knew their actual grades for each course they evaluated since the evaluations were completed after the examinations for the course.

An interesting question that has much to do with the relationship between grading and SET, concerns the *examination period* in which the grade for a course was received. In the Flemish higher education system, students who register for a course are given two chances to pass for (the examinations for) this course within the same academic year. If he/she fails for this course at the end of the semester, there is a second chance at the end of the academic year (mostly in the months August or September). The hypothesis then would be that students who pass their exam for the course in the first exam period will be more likely to give positive evaluations of the teacher and the course, no matter which

of the above hypotheses is true. We do not know about any study wherein this relationship between examination period and SET was included. In the present study, this relationship will be examined.

Another student characteristic that has not been studied that much, is student's *overall grade* by means of his/her overall grade based on all courses he/she registered for during one academic year. In line with the validity hypothesis (and thus in some way in contradiction with the grading leniency hypothesis), one could expect that better students (with higher overall grades) will give higher ratings to their teachers. With respect to this matter, Cohen (1981) found that students who gave the highest ratings to their instructors, were the ones who performed best on a common final exam. This suggests a strong connection between learning and SET. As already stated concerning the examination period in which the course grade was given, a similar argumentation can be made with respect to the *examination period wherein the overall grade was awarded*. An overall grade received in the second examination period makes it likely that a student had to retake examinations after all, which makes it worthy to include the examination period wherein the student's overall grade was awarded as a predictor for SET as well.

It seems plausible that students who attend classes regularly are more interested, more motivated, and maybe more convinced of their teacher's educational activities and teaching skills (Devadoss & Foltz, 1996; Fjortoft, 2005). Recent findings however show both statistically significant positive (Beran & Violato, 2005; Davidovitch & Soen, 2006) and non-significant (Guinn & Vincent, 2006) relationships between *class attendance* and SET. Probably these conflicting findings are partly due to differences in research design. The same conclusion can be drawn concerning the relationship between *student's gender* and SET (Centra & Gaubatz, 2000). Recent research however shows that there might be an interaction effect (or better: a gender pattern) between both student's and teacher's gender with respect to SET. Male students tend to choose less often than expected a female teacher as their best teacher (Basow, Phelan, & Capotosto, 2006). With respect to the last student characteristic variable in the present study, *student's age*, hardly any research has been done thus far. It is therefore not clear whether higher student ratings in upper-level courses are the result of the more specialized (and thus more interesting) subject-matter, or the more mature respondents (Wachtel, 1998).

Teacher characteristics

In a literature review on this topic, Feldman (1993) found no or only little statistically significant correlations (in favour of women, $r = .02$) between *teacher's gender* and SET. The few studies reporting statistically significant findings show that female teachers receive more favourable SET than their male colleagues (Griffin, 2004). However, as mentioned above, there might be interaction effects indicating that students tend to give higher ratings to same-gender teachers (Wachtel, 1998). The same can be said about the association between *teacher's rank* and SET. Most studies show no statistically significant linear correlations between rank and SET (Ting, 2000) and when significant effects are found, teachers with a higher rank receive higher ratings (Feldman, 1983). Concerning the relationship between *teacher's age* and SET, the majority of studies again shows no statistically significant correlations. In the few studies that do report a significant correlation, an inverse effect is found (Wachtel, 1998): teachers of greater age receive less favourable ratings (although all studies were cross-sectional).

Course characteristics

Feldman (in Marsh, 1987) and Theall and Franklin (2001) suggest that natural sciences courses are given lower SET-scores

Download English Version:

<https://daneshyari.com/en/article/372772>

Download Persian Version:

<https://daneshyari.com/article/372772>

[Daneshyari.com](https://daneshyari.com)