



Instructions for authors, subscriptions and further details:

<http://remie.hipatiapress.com>

An Exploratory Analysis of Major Dropdowns in Student Evaluation of Teaching Ratings in Higher Education

Santiago Leguey¹, Ana I. Cid-Cid¹, Rocío Guede-Cid¹, Jaime Prieto¹

1) Universidad Rey Juan Carlos

Date of publication: February 12th, 2023

Edition period: February 2023 – June 2023

To cite this article: Leguey, S., Cid-Cid, A. I., Guede-Cid, R., & Prieto, J. (2023). An Exploratory Analysis of Major Dropdowns in Student Evaluation of Teaching Ratings in Higher Education. *Multidisciplinary Journal of Educational Research*, 13(1), 91-113. <http://dx.doi.org/10.447/remie.10419>

To link this article: <http://dx.doi.org/10.447/remie.10419>

PLEASE SCROLL DOWN FOR ARTICLE

The terms and conditions of use are related to the Open Journal System and to [Creative Commons Attribution License](#) (CC-BY).

An Exploratory Analysis of Major Dropdowns in Student Evaluation of Teaching Ratings in Higher Education

Santiago Leguey
Rey Juan Carlos University

Ana I. Cid-Cid
Rey Juan Carlos University

Rocío Guede-Cid
Rey Juan Carlos University

Jaime Prieto
Rey Juan Carlos University

(Received: 17th May 2022; Accepted: 24th January 2023; Published: 12th February 2023)

Abstract

We present an exploratory study of Student Evaluation of Teaching (SET) surveys that focused on analyzing the major dropdowns that sometimes occur in SET scores of those university teachers who teach the same subject for at least two consecutive years. Specifically, the purpose of this study was to identify those SET survey items associated with major dropdowns in teacher evaluations. The sample comprised SET surveys ratings from one Spanish University over five years, for a total of 2,284 different teacher-subject pairs under analysis. The 5% percentile of the differences between the teacher and subject SET ratings in two consecutive years was used as the cut-off point to identify major dropdowns. Frequency tables were used for analysis. The results highlighted SET survey items related to teaching methodology as those most associated with major dropdowns in teachers' ratings. In particular, the item concerning teachers' compliance with class schedules showed the highest frequency of occurrence. Beyond the specific scores obtained or the specific items in which major dropdowns may take place, in this study we discuss the importance of incorporating major dropdowns analysis to SET surveys reports in the future.

Keywords: student evaluation of teaching, teacher evaluation, teaching quality, teaching practice, higher education.

Análisis Exploratorio de los Grandes Descensos en las Valoraciones Docentes del Profesorado en la Educación Superior

Santiago Leguey
Universidad Rey Juan Carlos

Ana I. Cid-Cid
Universidad Rey Juan Carlos

Rocío Guede-Cid
Universidad Rey Juan Carlos

Jaime Prieto
Universidad Rey Juan Carlos

(Recibido: 17 Mayo 2022; Aceptado: 24 Enero 2023; Publicado: 12 Febrero 2023)

Resumen

Presentamos un estudio exploratorio de las valoraciones docentes del profesorado enfocado al análisis de los grandes descensos que en ocasiones se producen en las puntuaciones que obtienen los profesores universitarios que imparten una misma asignatura en al menos dos cursos consecutivos. Específicamente, el propósito fue identificar aquellos ítems de las encuestas de valoración docente asociados a los grandes descensos de puntuación. Compusieron la muestra las encuestas de valoración docente de una universidad española durante 5 años, con 2.284 pares de profesor-asignatura diferentes. Se empleó el percentil 5% de las diferencias entre las puntuaciones de las encuestas de un profesor y asignatura en dos años consecutivos como punto de corte para identificar los descensos más pronunciados. Se emplearon tablas de frecuencias para el análisis. Los resultados destacaron los ítems relacionados con la metodología docente como los más asociados a los grandes descensos de puntuación. El ítem relativo al cumplimiento de los horarios de clase presentó la mayor frecuencia de ocurrencia. Más allá de las puntuaciones específicas obtenidas o de los ítems específicos en los que pueden ocurrir los mayores descensos, discutimos la importancia de incorporar su análisis a los informes de las encuestas de valoración docente del profesorado universitario.

Palabras clave: valoraciones docentes del profesorado, evaluación del profesorado, calidad de la docencia, práctica docente, educación superior

Within the current globalised higher education landscape, the use of rankings that measure and compare institutions' performance has acquired a very prominent role for universities wishing to attract new students, both in the national and the international context (Komotar, 2019). These rankings consider different performance indicators in regards to educational, research or business excellence and are considered a key measure for the comparison of higher education institutions (Hosier & Hoolash, 2017). In particular, the use of quality of education measures to assess academics' quality of teaching helps to depict the quality of the learning environment offered by institutions to their students (Falch et al., 2022). Quality of education indices are also considered key mechanisms for internal quality-assurance within the auditing and accreditation processes of the higher education institutions and programmes by establishing national and international ranking parameters that can be compared among universities (Fernandes & Singh, 2022).

Literature Review

During the last decades, different methods, procedures and instruments have been developed to evaluate the quality of teaching-learning processes at the university context, being the student evaluation of teaching (SET) surveys the most commonly adopted instruments to measure teaching performance in higher education (ASA, 2019). SET surveys ask students to rate instructors in various components related to effective teaching as an internal process of quality control for universities. Therefore, they allow institutions to collect valuable feedback on their teaching staff performance with a cheap and easy-to-implement procedure that usually takes the form of completion of anonymous surveys (Constantinou & Wijnen-Meijer, 2022). The information gathered through SET surveys is generally used by institutions in two ways. In its origins, it was mainly intended for formative purposes, providing diagnostic feedback to teachers on the quality of their teaching in order to facilitate their professional development (Spooren et al., 2017). However, and over the years, new purposes emerged concerning the use of SETs, such as its use for decision-making about faculty personnel when used summatively as an indicator of mastery (e.g., decisions about personnel

hiring, renewal or promotion), as relevant information for students to consult prospectively when deciding which course to enrol or which teacher to choose, or as a key factor for quality assurance in the accreditation of programs and institutions within the higher education context (ASA, 2019; Mart, 2017).

Typically, SET feedback is obtained through Likert-scale questionnaires surveying students' opinions on teaching quality. Despite criticisms surrounding whether students' ratings constitute a measure of students' satisfaction with the learning experience as consumers of a service rather than an indicator of teaching quality (Asare-Nuamah, 2017), or potential problems when considering these ratings as critical factors of judgment for personnel renewal, tenure or promotion (Uttl et al., 2017), SET surveys are widely considered to produce reliable results (Clayson, 2018; Feistauer & Richter, 2017), especially when administrators and institutions have access to research-based information about how to interpret and use the results (Linse, 2017).

Regarding the design of the instruments themselves, there are institutions (the fewest) that decide to develop their own SET instruments (generally through their quality departments) (e.g. Lizasoain-Hernández et al., 2017; Sanitya & Sinjindawong, 2017), whilst others (the most) choose to use standardized SET instruments. In this regard, the standardised SET instruments that are most widespread in the literature are the Questionnaire for Student Evaluation of Teaching SET37, the Student Evaluation of Educational Quality SEEQ, and the Students' Evaluation of Teaching Effectiveness Rating Scale SETERS (Moreno-Murcia et al., 2015).

However, despite this heterogeneity, there is a general consensus among policymakers, academics and researchers that underlines the need for the SET instruments to address the multidimensional nature of the teaching process in their design, thus needing to incorporate items that assess the different dimensions related to teaching performance (Nasser-Abu, 2017; Petek & Pope, 2022).

In this sense, in a relatively recent review of the literature, Bedggood and Donovan (2012) identified three as the dimensions most commonly measured in SET procedures: (i) quality of instruction, referring to both teachers skills and abilities, as well as to their kindness, approachability and enthusiasm;

(ii) task difficulty, in terms of the effort that students require to obtain the desired result; and (iii) academic development and stimulation, concerning how motivated and stimulated students feel and to what extent they believe they are improving their academic skills. Very interestingly, after conducting the review the authors highlighted the fact that “no single existing measure captures each of these constructs with depth and in the one instrument” (p.831).

SET surveys are administered to the students at the end of the academic period of each subject or course, with teachers generally having access to their SET results at the end of the semester. These results provide instructors with valuable information on how their students have perceived their performance during the course within each of the analyzed dimensions, constituting a measure that generally has a positive acceptance by teachers to improve the quality of their instruction in the upcoming courses (Jiang & Xiong, 2021). In this regard, it is important to highlight the importance of considering the potential biasing factors that can affect SET scores when interpreting the results.

In a relatively recent state-of-the-art review, Spooren et al. (2013) provide an overview of published studies that address student-related (e.g., students’ cognitive background, class attendance, students’ effort, students’ gender, students’ age, pre-course motivation, etc.), teacher-related (e.g., instructors’ teaching experience, instructors’ reputation, instructors’ age, instructors’ gender, instructors’ race, instructors’ rank, etc.), and course-related characteristics (e.g., class size, class attendance rate, class heterogeneity, course difficulty, course level, course workload, etc.) that might affect SET. In this sense, there are several studies that have used SET datasets collected from different universities worldwide to assess how these elements of bias may affect teaching evaluations (e.g. Boring, 2017; Mengel et al., 2019; Mitchell & Martin, 2018).

Regarding the methodological approach of SET research, published studies have used different methods and statistical techniques to examine SET results, focusing primarily on the reliability (i.e. the generalizability of ratings over students in the same class) and validity (i.e. are SETs valid as objective measures of teaching quality?) of student opinions (e.g., Clayson, 2018; Dennis, 2022; Kāhala & Thetsane, 2021), as well as their possible

relationship with different biasing factors such as gender stereotypes (e.g., Maričić et al., 2019; Renström et al., 2021), class size (e.g. Uttl et al., 2018), course workload (e.g. Ferrando et al., 2017), course difficulty (e.g. Griffin, 2016; Joyce, 2017), or the instructor motivation and effects of incentives (e.g. Alvero et al., 2019). Regarding the data analysis techniques used, these include, for example, correlation analyses to examine the relationships between the evaluations of different teachers in different courses and the relative influence of the teacher and the course on the ratings, or structural equation modelling (e.g., exploratory factor analysis, EFA, and confirmatory factor analysis, CFA) to identify the underlying relationships between the factors or variables included in SET surveys and thus be able to improve the design of the instruments (Castro-Morera et al., 2020, Spooren et al., 2017). However, an important limitation that many studies have presented is that they are based on samples collected from a single course (i.e., cross-sectional studies), therefore making it impossible to monitor changes in the ratings received by the same teacher over the years.

In this sense, long-term longitudinal studies, in which teachers' performance is monitored on many different occasions over a long period, offer an interesting approach to analyze systematic individual differences among teachers (i.e., teachers who systematically increase, reduce or maintain their performance stable) and to relate SET scores to variables such as teacher age, teaching experience, or faculty position (Sulis et al., 2019). In particular, the analysis of the ratings of the same teacher and the same subject in two consecutive years may provide valuable information on how that teacher's performance has changed with the near experience of one more year of teaching the subject.

Although the evaluating students will be different (i.e., individual characteristics, academic performance, in addition to possible differences in the number of students responding), teachers will be able to know the items in which their students' ratings improve, remain or get worse, helping them to identify their own strengths and weaknesses in teaching in the short term (Mart, 2017). Even though all this information may be of interest, focusing more explicitly on those components in which ratings significantly drop down is commonly perceived as the most important by teachers to update their courses in response to student feedback (Chan et al., 2014).

However, studies that have specifically analyzed major dropdowns in teachers' ratings and their possible interaction effects with students, teachers, or course characteristics are few and mainly focused on mean stability analysis within long-term longitudinal studies. Furthermore, the majority of studies that have conducted year-to-year and teacher-subject analyses of SET surveys emerged with comparison purposes as a result of the proliferation of Internet-based applications within the higher education context and the subsequent widespread change from traditional paper-based to computer web-based evaluation systems (e.g. Baldo et al., 2020; Kuch & Roberts, 2018; Mitchell & Morales, 2018; Stanny & Arruda, 2017; Treischl & Wolbring, 2017).

Within this context, we attempted to account for these shortcomings of previous research by presenting an exploratory study of SET surveys that focused on analyzing the major dropdowns that sometimes occur in SET scores of those university teachers who teach the same subject for at least two consecutive years. Specifically, the purpose of this study was to identify those SET survey items associated with major dropdowns in teacher evaluations of a particular subject and to analyze their immediate impact on next year's scores.

Method

Sample

The study considered five years of SET surveys of the Rey Juan Carlos University (Madrid, Spain). Specifically, the sample comprised all the SET surveys of those subjects taught by the same teacher for at least two consecutive years, for a total number of 2,284 different teacher-subject observations. Table 1 shows the distribution of teachers based on the number of consecutive years in which they taught the subjects comprised in the sample. The SET surveys of different classes of the same teacher and subject in one particular year were grouped for analysis purposes. All this yielded a total of 3,893 comparisons under analysis.

Table 1.

Distribution of teachers based on the number of consecutive years in which they taught the same subjects

No. of consecutive years	No. of teachers
2	1,268
3	576
4	302
5	138

Procedure and Instrument

Paper-based SET surveys were provided to students by a team of surveyors (administrative staff) in a face-to-face classroom setting according to standardized written instructions. Evaluations of single-term subjects were collected at the end of each academic term, between two and four weeks before the start of the examination period set up by the University. Evaluations of full-year subjects were collected at the end of the second term. Optical mark recognition software (Evaldara, Dara Group) was used to read and process the answers automatically.

The survey asked students to meet their satisfaction with the teacher of a given subject, with a five-point Likert scale questionnaire ranging from 1: “Strongly disagree” to 5: “Strongly agree”. The survey contained ten items and was the standard instrument used by the University during the research. Nine of the items were relative to a variety of factors related to teaching performance during the course.

A detailed analysis of the questionnaire allowed identifying three dimensions captured by the instrument, which referred to planning and organization of the subject, teacher obligations and teaching methodology. The tenth item retrieved the overall satisfaction of the student with the teacher (OS). Table 2 shows the captured dimensions and the detail of the items included in each dimension. The table includes identifications codes for each dimension and item that are subsequently used to display the results of the analysis.

Table 2.

Survey items and dimensions captured by the instrument.

Dimension	Survey items	Identification code
Planning and organization (PO)	The teacher explains in detail to the students the teaching guide of the subject at the beginning of the course	PO1
	The teacher has informed clearly about the assessment criteria of the subject	PO2
	The teacher, in addition to the face-to-face classes, has planned complementary activities (e.g., problem-solving, readings, practical exercises) that facilitate the learning of the subject	PO3
Teacher obligations (TO)	The teacher respects the class schedules	TO1
	The teacher is available to attend to the students	TO2
	Teaching activities meet the objectives, contents and methodology specified in the teaching guide of the subject	TO3
Teaching methodology (TM)	The teacher adequately clarifies the doubts of the different activities proposed in the subject	TM1
	The teacher explains clearly	TM2
	The development of the subject allows me adequate monitoring and learning	TM3
Overall satisfaction (OS)	Taking into account all the aspects mentioned, I am satisfied with the work carried out by the teacher	OS

Case Selection and Analysis

The study aimed to identify major dropdowns between teacher-subject ratings for those teachers who taught the same subject for at least two consecutive years. In particular, the analysis aimed to identify those SET

survey items associated with major dropdowns occurrence. For this purpose, we adhere to Fisher's argument that establishes the frequency of extraordinary events at 5% (Fisher, 1970) to set the 5% percentile of the differences between the teacher and subject SET ratings in two consecutive years as the cut-off point to identify major dropdowns. All teachers' SET ratings in which at least one of the items experienced a major dropdown, or the average score of all the items did so, were considered. A major dropdown was considered isolated when there was no major average dropdown, but there was a major dropdown in at least one of the survey items. Frequency tables were calculated to show the relationship between major dropdowns and survey items.

The research complied with the ethical principles of research of the university where the research was conducted and with the Ethical Guidelines for Educational Research published by the British Educational Research Association (BERA, 2018).

Results

Percentiles

Table 3 shows the values from which 1%, 5%, 10%, 25% and 50% of the differences in the teacher ratings in consecutive years were found (percentiles). Data are presented for each item of the survey, as well as for the set of ten items and for the mean of all the items. The 5% percentile of the lowest differences between the survey items was between -0.78 and -0.97. The 1,963 items that exceeded this limit corresponded to 574 different pairs of teacher-subject ratings, of which 194 were major average dropdowns. In addition, the 10% percentile of the lowest differences between the survey items was between 0.56 and 0.71.

These values are relatively high if we take into account that the amplitude of the scale is only 4 points. In other words, we find that the magnitude of the variations under analysis is considerable.

Table 3.

Percentiles of differences in the teacher ratings in consecutive years.

Survey item	P1%	P5%	P10%	P25%	P50%
PO1	-1.26	-0.81	-0.56	-0.25	0.06
PO2	-1.35	-0.78	-0.57	-0.24	0.06
PO3	-1.30	-0.83	-0.59	-0.28	0.04
TO1	-1.56	-0.91	-0.64	-0.29	0.03
TO2	-1.38	-0.83	-0.61	-0.29	0.02
TO3	-1.40	-0.89	-0.66	-0.33	-0.01
TM1	-1.46	-0.92	-0.65	-0.30	0.03
TM2	-1.44	-0.97	-0.69	-0.31	0.03
TM3	-1.39	-0.88	-0.65	-0.27	0.07
OS	-1.49	-0.94	-0.71	-0.33	0.02
Set of all items	-1.41	-0.87	-0.63	-0.29	0.03
Mean of all items	-1.23	-0.78	-0.56	-0.25	0.03

Frequencies

Table 4 shows the percentage of the survey items in which there was a major dropdown by type. The lack of respect for class schedules (TO1) was the item with the highest frequency among the items in which isolated major dropdowns occurred (28%). However, it was the item with less frequency of occurrence among the items with major average dropdowns (54%). Concerning major average dropdowns, all items showed frequencies of occurrence greater than 50%.

The average dropdowns were unequivocally manifested in the overall satisfaction of the student with the teacher (the tenth item of the survey, showing a major dropdown in 82% of cases) and in the items related to the teaching methodology (between 76% and 79%). In this case, all the items had a relative frequency of occurrence greater than 50%.

Table 4.

Percentage of items in which there is a major dropdown by type.

Survey item	Isolated major dropdowns	Major average dropdowns	All major dropdowns
PO1	11%	57%	26%
PO2	10%	57%	26%
PO3	15%	58%	30%
TO1	28%	54%	37%
TO2	11%	67%	30%
TO3	19%	67%	35%
TM1	18%	79%	39%
TM2	23%	79%	42%
TM3	13%	76%	35%
OS	23%	82%	43%

Discussion

This study presented an exploratory study of SET surveys ratings that focused on analyzing major dropdowns in SET ratings for those teachers who taught the same subject for at least two consecutive years. The analysis aimed to identify those SET survey items associated with major dropdowns occurrence. For that purpose, the case selection and analysis procedure set that, while major average dropdowns reflect students' lower overall satisfaction levels compared to the previous year, an isolated major dropdown indicates that the satisfaction with that particular item has decreased compared to last year and that that item is standing out from the rest of the survey items.

In general terms, major average dropdowns in teachers' ratings were associated with major drops in most of the items of the SET survey. This suggests that students' satisfaction with their teachers does not depend on a single aspect of teaching, but rather is related to multiple items and dimensions. This would contribute to providing further evidence to support

the multidimensionality of teaching and how this affects the instructors' general teaching skills and competencies (Nasser-Abu, 2017). In this regard, previous research has highlighted the complex and multifaceted nature of teaching and learning processes, and therefore, the multifactorial component of the SET measures (Petek & Pope, 2022). Specifically, the results of the present study pointed out various aspects of teaching methodology as those most associated with major dropdowns in teachers' ratings. In particular, relative to three of the items of the survey (i.e., TM1: 'The teacher adequately clarifies the doubts of the different activities proposed in the subject', TM2: 'The teacher explains clearly'; TM3: 'The development of the subject allows me adequate monitoring and learning'). These results would be in line with a previous study that specifically found major dropdowns in the factors related to organization and clarity when evaluations changed from a traditional paper-based context to an online computer-based setting (Mau & Opengart, 2012). However, a detailed analysis of the questionnaire used by the University in Mau's research (e.g., four dimensions, 20 items, four-point Likert scale, no overall satisfaction item) reveals its many differences with the one used in the present study, so the need to be cautious when comparing the results should be emphasized.

In fact, one of the main difficulties that has been identified in the literature when discussing the results of SET studies is the lack of homogeneity in the questionnaires (Ching et al., 2018). From its origins, multiple scales and questionnaires have been developed by institutions and academics to address the dimensionality of student satisfaction and teaching effectiveness (Spooren et al., 2017). Although there are questionnaires that have had greater prominence in published research (e.g., Student Evaluation of Educational Quality (SEEQ), SET37, Students Evaluation of Teaching Effectiveness Rating Scale, Teaching Proficiency Item Pool; for a brief overview, see Moreno-Murcia et al., 2015), currently there is no international consensus on the use of a single questionnaire, and it seems difficult that it will exist given the particular characteristics of each country, educational system, university, etc. In this particular sense, interesting projects that seek to identify good practices of SET related processes among higher education institutions with the aim of optimizing the assessment of course quality and teaching effectiveness, continue to be carried out by institutions (e.g., the

SET Project: Student Evaluations of Teaching, Measuring and Enhancing Course Quality and Teaching Quality, conducted by the Rotterdam School of Management to explore student evaluation practices at business universities and institutions across Europe, see [Scheepers, 2019](#)).

Returning to the results of the present study and focusing now on the isolated major dropdowns, these were mainly associated with either one or two of the items of the SET survey. In particular, the item concerning teachers' compliance with class schedules showed the highest frequency of occurrence when the dropdown was manifested in a single item of the SET survey. Specifically, it appeared three times more frequently than the next item did (the one referred to teaching activities meeting the objectives, contents and methodology specified in the teaching guide of the subject). It was also the item with the highest frequency when isolated dropdowns were manifested in two of the survey items, mainly combined with the item relative to the overall (dis)satisfaction and with the one related to the adequacy of the teaching activities with the teaching guide of the subject. This could suggest that compliance with class schedules is an aspect highly valued by students and that it can penalize teachers who do not respect them, even though students are satisfied with their teaching. This is in line with a published study focused on the development and validation of a new SET questionnaire in a Spanish university (University of Jaen; [Molero & Ruiz, 2005](#)), in which the compliance with class schedules was the item with the highest score when students (n=15,291) were asked to evaluate the variables they considered most important when assessing instructors' quality of teaching. In this regard, in [Casero \(2008\)](#) work, it is observed how the item related to the fulfilment of class schedules and teaching obligations is present in several SET questionnaires of Spanish universities. However, the author proposes its suppression because these are part of the teachers' duties and not aspects of teaching competence. In this sense, it should be noted that the item related to schedule compliance by the teacher is not included in the majority of the most prominent SET questionnaires in the published literature (e.g., SEEQ; SET37; Exemplary Teacher Course Questionnaire ETCQ).

Once the items associated with major dropdowns have been identified, one critical issue is establishing whether the observed effects are due exclusively to random variations or are due to other causes. Nonetheless,

given that major differences may be due to exceptionally high SET ratings in the reference course, or exceptionally low ones in the course in which the dropdown is detected, scores closer to the mean (average) could be found the year after the dropdown without teacher's action or causal modification taking place. In this sense, previous research has established relationships between factors that are not strictly educational and SET ratings, such as teacher's reputation or personality. For example, Griffin (2016) analyzed a sample of SET surveys from 914 students enrolled in 47 randomly selected courses at a regional American university over a period of five semesters, and found that students who heard positive information about the teacher prior to enrolling in the course (i.e., positive reputation), rated both the instructor and the course higher than those students who heard negative information about the teacher (i.e., negative reputation). Within this same line of research, Kim and MacCann (2018) analyzed a sample of 515 students of first year and 45 teachers. Very interestingly, the authors reported five teacher personality-related domains (i.e., openness, conscientiousness, extraversion, agreeableness, and emotional stability) as positively associated with SET scores in terms of teacher effectiveness. Moreover, published research has found strong relationships between students' perception of teacher's personality or charisma and teaching effectiveness ratings (Huang & Lin, 2014). However, these aspects (personality and reputation) are generally not susceptible to abrupt variations over time, so their possible impact on the score differences would be of little relevance. Furthermore, whether these or other aspects could affect students' general perception of their satisfaction with a teacher, these shall affect the whole set of scores and, therefore, the overall teacher evaluation (Kim et al., 2019).

Finally, regarding the role of the teachers towards SET ratings once results are reported to them, several questions of interest arise. Are teachers aware of major dropdowns occurrence? Do they consider major dropdowns to be important? And if so, do they take any action to refine their teaching? Teachers usually have access to their subjects SET ratings reports after the end of the course, providing them with valuable insights into student feedback. In most cases, these annual reports are based on the scores derived from Likert-type SET surveys of each subject for each teacher. However, SET survey reports of results do not usually show a history of results of past

years, in some cases due to the impossibility of establishing comparisons (e.g. a teacher who begins to teach a new subject), but in many others, where comparisons are possible, due to their generalized conception as annual reports on the performance of a teacher in a specific year in a specific subject. Hence, the information that the teacher can use to refine his/her teaching is limited to what happened last year, leaving it up to his/her own initiative and interest to consult the reports of previous years to establish comparisons in order to identify the items or dimensions in which his/her performance worsens, is maintained, or improves over the years. This fact can cause the taking of specific actions by teachers to be seriously affected, suggesting, in this case, the incorporation of major dropdowns analysis to SET survey reports to provide teachers with valuable insights into the stability or change of their teaching effectiveness over time (Sulis et al., 2019). In this regard, it is important to highlight the importance of considering the potential biasing factors that can affect SET scores when interpreting the results of a particular course (e.g. class size, class attendance rate, class heterogeneity, instructors' teaching experience; Feistauer & Richter, 2018). Hence, monitoring teachers' performance over different years can better deal with these potential biasing factors.

Conclusion

The present research presented an exploratory analysis of those major dropdowns in SET scores that sometimes affect university teachers who teach the same subject in consecutive years. Our study provides evidence that SET survey items related to teaching methodology are the those most associated with major dropdowns in teachers' ratings. Specifically, the item concerning teachers' compliance with class schedules showed the highest frequency of occurrence. From a broader view, beyond the specific scores obtained and/or the specific items in which these major dropdowns may occur, our research suggests that incorporating major dropdowns analysis to SET reports could provide teachers with relevant information to identify their major weaknesses in their teaching in the short term. This might help instructors to improve and enhance the quality of their teaching in subsequent courses, especially when combined with appropriate communication (e.g.,

expert consultants to assist the teachers in the interpretation of SET results), institutions support (e.g., department heads leading campaigns for better instructions and providing incentives for improving teaching quality), and self-reflection. The evaluation of teaching is a complex process in which multiple factors are involved. Large university institutions can hardly attend SET results in detail, but perhaps they can handle those few big differences that tell us what students' value most about their teachers and what causes their dissatisfaction. Overall, carrying out new studies that gather the teachers' opinions involved in major dropdowns in SET ratings might be of great interest to better understand the behaviour of the scores and the possible reactions of the teaching staff in subsequent years. These studies shall include qualitative research methods (e.g., small-group discussions, semi-structured interviews, in-depth interviews) to investigate teachers' beliefs, attitudes and reflections on their experiences with SET survey ratings when experimenting major dropdowns (e.g., consideration given to the results, alleged causes of the dropdowns, measures adopted by the teachers).

References

- Alvero, A. M., Mangiapanello, K., & Valad, J. (2019). The effects of incentives, instructor motivation and feedback strategies on faculty evaluation response rates in large and small class sizes. *Assessment & Evaluation in Higher Education*, 44(4), 501–515. <https://doi.org/10.1080/02602938.2018.1521913>.
- ASA, American Sociological Association (2019). *Statement on Student Evaluation of Teaching*. https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_sept52019.pdf
- Asare-Nuamah, P. (2017). International Students' Satisfaction: Assessing the Determinants of Satisfaction. *Higher Education for the Future*, 4(1), 44–59. <https://doi.org/10.1177/2347631116681213>
- Baldo, C. M., Snyder, J., & Holguin, A. (2020). Revisiting the Online versus Face-to-Face teaching evaluations. *International Journal of Education and Development using Information and Communication Technology*, 16(2), 144–157.

- Bedggood, R. E., & Donovan, J. D. (2012). University performance evaluations: what are we really measuring? *Studies in Higher Education, 37*(7), 825–842.
<https://doi.org/10.1080/03075079.2010.549221>
- Boring, A. (2017). Gender biases in student evaluations of teachers. *Journal of Public Economics, 145*, 27–41.
<https://doi.org/10.1016/j.jpubeco.2016.11.006>
- BERA, British Educational Research Association (2018). *Ethical Guidelines for Educational Research*. London: BERA.
- Casero, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. [Proposal for a consensual questionnaire for evaluating the quality of university among students and teachers]. *Revista de Investigación Educativa, 26*(1), 25–44.
- Castro-Morera, M.; Navarro-Asencio, E. & Blanco-Blanco, Á. (2020). The quality of teaching as perceived by students and university teachers: analysis of the dimensionality of a teacher evaluation questionnaire. *Educación XXI, 23*(2), 41–65.
<https://doi.org/10.5944/educXXI.25711>
- Chan, C. K. Y., Luk, L. Y. Y., & Zeng, M. (2014). Teachers' perceptions of student evaluations of teaching. *Educational Research and Evaluation, 20*(4), 275–289.
<https://doi.org/10.1080/13803611.2014.932698>
- Ching, G. (2018). A literature review on the student evaluation of teaching: An examination of the search, experience, and credence qualities of SET. *Higher Education Evaluation and Development, 12*(2), 63–84.
<https://doi.org/10.1108/HEED-04-2018-0009>
- Clayson, D. E. (2018). Student evaluation of teaching and matters of reliability. *Assessment & Evaluation in Higher Education, 43*(4) 666–681. <https://doi.org/10.1080/02602938.2017.1393495>
- Constantinou, C., & Wijnen-Meijer, M. (2022). Student evaluations of teaching and the development of a comprehensive measure of teaching effectiveness for medical schools. *BMC Medical Education, 22*, Article 113. <https://doi.org/10.1186/s12909-022-03148-6>
- Dennis, C. (2022). The student evaluation of teaching and likability: what the evaluations actually measure. *Assessment & Evaluation in Higher Education, 47*(2), 313–326.
<https://doi.org/10.1080/02602938.2021.1909702>

- Falch, T., Iversen, J. M. V., Nyhus, O. H., & Strøma, B. (2022). Quality measures in higher education: Norwegian evidence. *Economics of Education Review*, 87, Article 102235. <https://doi.org/10.1016/j.econedurev.2022.102235>
- Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, 42(8), 1263–1279. <https://doi.org/10.1080/02602938.2016.1261083>
- Feistauer, D., & Richter, T. (2018). Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. *Studies in Educational Evaluation*, 59, 168–178. <https://doi.org/10.1016/j.stueduc.2018.07.009>
- Fernandes, J. O., & Singh, B. (2022). Accreditation and ranking of higher education institutions (HEIs): review, observations and recommendations for the Indian higher education system. *The TQM Journal*, 34(5), 1013–1038. <https://doi.org/10.1108/TQM-04-2021-0115>
- Ferrando, A., Martínez-Tomás, M. C., & Cruz, J. L. (2017). Student Evaluation of Teaching (SET): analysis of the workload in the 3th year of the physics degree at the University of Valencia. *EDULEARN17 Proceedings*, pp. 5289–5294.. <https://dx.doi.org/10.21125/edulearn.2017.2192>
- Fisher, R. A. (1970). *Statistical Methods for Research Workers*. Hafner.
- Griffin, B. W. (2016). Perceived autonomy support, intrinsic motivation, and student ratings of instruction. *Studies in Educational Evaluation*, 51, 116–125. <https://doi.org/10.1016/j.stueduc.2016.10.007>
- Hosier, M., & Hoolash, B. K. A. (2017). The effect of methodological variations on university rankings and associated decision-making and policy. *Studies in Higher Education*, 44(1), 201–214. <https://doi.org/10.1080/03075079.2017.1356282>
- Huang, Y. C., & Lin, S. H. (2014). Assessment of Charisma as a Factor in Effective Teaching. *Educational Technology & Society*, 17(2), 284–295.
- Jiang, H. C., & Xiong, Y. (2021). The indicators, characteristics and enlightenment of students' evaluation of teaching in Japanese national universities: taking education as an example. *Higher Education Research*, 42, 103–109.

- Joyce, A. (2017). Course Difficulty and Its Association with Student Perceptions of Teaching and Learning-RESEARCH. *Kentucky Journal of Excellence in College Teaching and Learning*, 14, Article 4.
- Kāhala, P., & Thetsane, R. (2021). The Validity and Reliability of Student Evaluation of Teaching at the National University of Lesotho. *International Journal of African Higher Education*, 8(1), 139–157. <https://doi.org/10.6017/ijahe.v8i1.13367>
- Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A Meta-Analysis of the Effects of Teacher Personality on Teacher Effectiveness and Burnout. *Educational Psychology Review*, 31, 163–195. <https://doi.org/10.1007/s10648-018-9458-2>
- Kim, L., & MacCann, C. (2018). Instructor personality matters for student evaluations: Evidence from two subject areas at university. *British Journal of Educational Psychology*, 88(4), 584–605. <https://doi.org/10.1111/bjep.12205>
- Komotar, M. H. (2019). Global university rankings and their impact on the internationalisation of higher education. *European Journal of Education*, 54(2), 299–310. <https://doi.org/10.1111/ejed.12332>
- Kuch, F., & Roberts, R. M. (2018). Electronic in-class course evaluations and future directions. *Assessment & Evaluation in Higher Education*, 44(5), 726–731. <https://doi.org/10.1080/02602938.2018.1532491>.
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94–106. <https://doi.org/10.1016/j.stueduc.2016.12.004>
- Lizasoain-Hernández, L.; Etxeberria-Murgiondo, J., & Lukas-Mujika, J. F. (2017). A proposal for a new questionnaire for the evaluation of teachers at the University of the Basque Country. Dimensional, differential and psychometric study. *RELIEVE*, 23(2), Article 1. <https://doi.org/10.7203/relieve.23.2.10436>
- Maričić, M., Đoković, A., & Jeremić, V. (2019). The Validity of Student Evaluation of Teaching: Is There a Gender Bias? *Croatian Journal of Education*, 21(3), 743–775. <https://doi.org/10.15516/cje.v21i3.3177>
- Mart, C. T. (2017). Student Evaluations of Teaching Effectiveness in Higher Education. *International Journal of Academic Research in Business and Social Sciences*, 7(10), 57–61. <http://dx.doi.org/10.6007/IJARBS/v7-i10/3358>

- Mau, R. R., & Opengart, R. (2012). Comparing Ratings: In-class (paper) vs. Out of Class (online) Student Evaluations. *Higher Education Studies*, 2(3), 55–68. <https://doi.org/10.5539/hes.v2n3p55>
- Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535–566. <https://doi.org/10.1093/jeea/jvx057>
- Mitchell, K., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652. <https://doi.org/10.1017/S104909651800001X>
- Mitchell, O., & Morales, M. (2018). The effect of switching to mandatory online course assessments on response rates and course ratings. *Assessment and Evaluation in Higher Education*, 43(4), 629–639. <https://doi.org/10.1080/02602938.2017.1390062>
- Molero, D., & Ruiz, J. (2005). La evaluación de la docencia universitaria. Dimensiones y variables más relevantes. [The evaluation of university teaching. Dimensions and most relevant variables]. *Revista de Investigación Educativa*, 23(1), 57–84.
- Moreno-Murcia, J. A., Silveira, Y., & Belando, N. (2015). Questionnaire Evaluating Teaching Competencies in the University Environment. Evaluation of Teaching Competencies in the University. *Journal of New Approaches in Educational Research*, 4(1), 54–61.
- Nasser-Abu, F. (2017). Teaching in higher education: Good teaching through students' lens. *Studies in Educational Evaluation*, 54, 4–12. <https://doi.org/10.1016/j.stueduc.2016.10.006>
- Petek, N., & Pope, N. G. (2022). The multidimensional impact of teachers on students. *Journal of Political Economy*. Advance online publication. <https://doi.org/10.1086/722227>
- Renström, E. A., Gustafsson-Sendén, M., & Lindqvist, A. (2021). Gender Stereotypes in Student Evaluations of Teaching. *Frontiers in Education*, 5, Article 571287. <https://doi.org/10.3389/educ.2020.571287>
- Sanitya, R., & Sinjindawong, S. (2017). Design of Students' Evaluation of Teaching: The Case of Sripatum University. *International Journal of Information and Education Technology*, 7(5), 367–373. <https://doi.org/10.18178/ijiet.2017.7.5.896>
- Scheepers, A. (2019). *SET Project: Student Evaluations of Teaching, Measuring and Enhancing Course Quality and Teaching Quality*.

Rotterdam School of Management. https://equal.network/wp-content/uploads/2019/03/SET_Ad-Scheepers-.pdf

- Spooren, P., Brockx, B. & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>
- Spooren, P., Vandermoere, F., Vanderstraeten, R., & Pepermans, K. (2017). Exploring high impact scholarship in research on student's evaluation of teaching (SET). *Educational Research Review*, 22, 129–141. <https://doi.org/10.1016/j.edurev.2017.09.001>
- Stanny, C. J., & Arruda, J. E. (2017). A comparison of student evaluations of teaching with online and paper-based administration. *Scholarship of Teaching and Learning in Psychology*, 3(3), 198–207. <https://doi.org/10.1037/stl0000087>
- Sulis, I., Porcu, M., & Capursi, V. (2019). On the Use of Student Evaluation of Teaching: A Longitudinal Analysis Combining Measurement Issues and Implications of the Exercise. *Social Indicators Research*, 142, 1305–1331. <https://doi.org/10.1007/s11205-018-1946-8>
- Treischl, E., & Wolbring, T. (2017). The Causal Effect of Survey Mode on Students' Evaluations of Teaching: Empirical Evidence from Three Field Experiments. *Research in Higher Education*, 58(8) 904–921. <https://doi.org/10.1007/s11162-017-9452-4>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54(1), 22–42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Uttl, B., Bell, S., & Banks, K. (2018). Student Evaluation of Teaching (SET) Ratings Depend on the Class Size: A Systematic Review. *Proceedings of International Academic Conferences*, 8110392. <https://doi.org/10.20472/IAC.2018.044.050>

Santiago Leguey is a senior lecturer on Mathematics and Statistics at the Rey Juan Carlos University (Spain). His research interests include applied statistics and education, with particular interest to the field of higher education. He is the Chief Manager of the University Center for Applied Social Studies CUESA.

ORCID: <http://orcid.org/0000-0002-2438-8455>

Ana I. Cid-Cid is a senior lecturer on Mathematics and Statistics at the Rey Juan Carlos University (Spain). She holds a PhD in Business and Economics by the Complutense University of Madrid. Her research interests focus on the assessment of both students' and teachers' performance in the higher education context.

ORCID: <http://orcid.org/0000-0001-9418-2113>

Rocío Guede-Cid is a lecturer on Mathematics and Statistics at the Rey Juan Carlos University (Spain). She holds a PhD by the Rey Juan Carlos University. Her research interests focus on the field of innovation and assessment in higher education, as well as on the analysis of the technology transfer process.

ORCID: <http://orcid.org/0000-0003-3450-615X>

Jaime Prieto is a lecturer on information and communications technology at the Rey Juan Carlos University (Spain). He holds a PhD by the Technical University of Madrid. His research interests focus on the field of ICT and education, with particular interest to the field of higher education.

ORCID: <http://orcid.org/0000-0002-4362-0889>

Contact Address: Facultad de Ciencias de la Economía y de la Empresa, Universidad Rey Juan Carlos, Paseo de los Artilleros, s/n, 28032, Madrid, Spain.

Email: jaime.prieto@urjc.es