

## ANALYSIS OF PROFESSORS' EVALUATION AT LA SALLE UNIVERSITY MÉXICO FROM 2010 TO 2016: WHAT THE RESULTS INDICATE?

Martin Flégl<sup>1✉</sup>, María Bertha Fortoul Ollivier<sup>2</sup>, Václav Švec<sup>3</sup>, Jennie Brand Barajas<sup>4</sup>, Christian Vizuet<sup>4</sup>

<sup>1✉</sup>Grupo de investigación Análisis de rendimiento y eficiencia de factores sociales, económicos y educativos, Facultad de Negocios, Universidad La Salle México, Ciudad de México, México, martin.flegl@hotmail.com

<sup>2</sup>Facultad de Humanidades y Ciencias Sociales, Universidad La Salle México, Ciudad de México, México

<sup>3</sup>Department of Management, Faculty of Economics and Management, Czech University of Life Sciences Prague, Prague, Czech Republic

<sup>4</sup>Coordinación de Formación Docente, Universidad La Salle México, Ciudad de México, México

### Highlights

- System of professors' evaluation regarding teaching quality
- Results indicate low discrimination ability in the evaluation
- Significant differences in the evaluation requiring additional analysis
- Proposed structural changes within the evaluation system

### Abstract

La Salle University México (La Salle) uses an internal system of professors' evaluation, which main purpose is to evaluate professors' performance and secure high quality of teaching at all of its faculties. Since its inception in 2010, La Salle has obtained 517,635 individual evaluations of 45,346 courses. However, no additional analysis of the obtained results has ever been done. This article provides introductory analysis of the accumulated results at faculty level. The main objective is to analyze whether there are differences between faculties regarding the evaluation. Although the results are highly skewed towards the maximal evaluation at all faculties, there are statistically significant differences. The next important task is to investigate what factors influence the evaluation. Moreover, as this is the introductory analysis, the article concludes with possible future steps that should be consider regarding eventual structural changes in the evaluation system.

### Keywords

Analysis of variance, Discrimination ability, Games-Howell test, Higher education, Professors' evaluation

### Article type

Full research paper

### Article history

Received: May 24, 2017

Received in revised form: September 4, 2017

Accepted: September 5, 2017

Available on-line: October 17, 2017

Flégl M., Fortoul Ollivier M.B., Švec V., Brand Barajas J., Vizuet Ch. (2017) "Analysis of professors' evaluation at La Salle University México from 2010 to 2016: What the results indicate?", *Journal on Efficiency and Responsibility in Education and Science*, Vol. 10, No. 3, pp. 76-85, online ISSN 1803-1617, printed ISSN 2336-2375, doi: 10.7160/eriesj.2017.100303.

### Introduction

Nowadays, many universities introduce internal evaluations systems in order to secure and improve education quality. For this purpose, the use of students' evaluations of professors has become very common and popular. The aim of these evaluations is to measure professors' performance and quality. In most of the cases, the internal evaluation systems use online questionnaires with diverse set of questions divided into several areas (dimensions). The most common areas are educative, didactic and pedagogic. However, an evaluation system can also consider areas such as social and ethics, or teaching and learning (Hein, Kroenke and Rodrigues Júnior, 2015). The choice of areas and included questions usually correspond to a university strategy and culture, as well as to the main purpose of the evaluation. Usually, questionnaires include open-ended questions where students can express their opinions and/or attach comments related to the evaluation.

University authorities uses the results of professors' evaluations to secure a constant development of a teaching quality. Moreover, the results are used to solve teaching related problems, for a motivation of teachers for their personal development, as well as for hiring and promotion decisions (Becker and Watts, 1999). In Mexico, it is common that from elementary school until high school, professors are evaluated by their authorities and not by students. From bachelor level, students begin evaluating professors, and it is their first experience with an evaluation. Therefore, students at lower level do not have possibility directly influence quality of teaching. As the professors' evaluation is commonly anonymous, the idea is to obtain valuable information from students who have direct contact with a professor. Hence, students must not be afraid to

evaluate their professors. In this case, students must perceive the evaluation as a chance to be directly involved in teaching quality improvements at university. However, some problems may arise regarding the reliability of the evaluation. As Braga, Paccagnella and Pellizzari (2014) point out, students' objectives might be different from the university authorities. "Students may simply care about their grades, whereas the university cares about their learning and the two might not be perfectly correlated, especially when the same professor is engaged both in teaching and in grading." (Braga, Paccagnella and Pellizzari, 2014: 72)

In addition, any evaluation must consider specifics related to professors' evaluation. If the authorities seek to receive valuable information regarding teaching, the evaluation system must attract students' interest and needs (Brand Barajas, 2014). What is more, professors' evaluations are biased by gender and attractiveness of a professor (Basow, 2000; Felton et al., 2008; Silva et al., 2008). Usually, male and attractive professor receives better evaluation than female professor. In a similar manner, positive and negative evaluation is directly linked to students' interest about a course and the ease of the course (Leung, Jiang and Busser, 2013). The higher the interest is, the better the evaluation is. On the other hand, the ease of a course can have both a positive impact, as well as negative impact on an evaluation (Felton et al., 2008, Marsh and Roche, 2000).

La Salle University México (La Salle) uses System of professors' evaluation (called SED 2.0) to evaluate professors' teaching quality at the end of each semester at bachelor level. The second version of the system was launched in 2010 and the first results of the evaluation were obtained the same year in December (related

to semester August-December). Nowadays, the authorities of La Salle have access to a valuable database containing more than 500,000 individual evaluations throughout 13 semesters of SED 2.0 history. Although this valuable database exists, no complex statistical analysis of the data has been provided. As consequence, no information about the relations within the evaluation exists. As the authorities seek to secure and improve teaching quality at La Salle, it is important to provide analysis of the current state of nature. Similarly, before starting changing the current evaluation system, it is necessary to analyze and understand current inter-relations within the system.

The objective of this article is to analyze whether significant differences in the evaluation exist between faculties at La Salle University México. For this purpose, we analyze results from the SED 2.0 using data from the period 2010 - 2016. Additional objective is to propose future recommendations regarding SED 2.0 based on the gained results.

The article is divided as follows: We begin with a brief introduction into the systems of professors' evaluations. In the following part, we describe evaluation system that is currently used at La Salle, describe the dataset, as well as methods used in the analysis. Further, we continue with detailed explanation of results, followed by a discussion over the achieved results. We conclude the article by some perspective for future research.

## Materials and Methods

### Evaluation system SED 2.0 (Sistema de Evaluación Docente)

Evaluation system SED 2.0 was created at La Salle University México in 2001. The purpose of the creation was a necessity to evaluate professors' performance in a fast, safe and reliable way according to the institutional philosophy centered on a person and his/her integral training. SED 2.0 is maintained by Department of Teacher Education (DPE), i.e. DPE maintains its design, development, results dissemination, etc.

The first version of SED, instrument for professors' evaluation by university students, was created in 2001. This system consisted of 30 questions, which were divided into 5 separate areas: self-evaluation (5 questions), interaction (8), intervention (10), professional (2), and administration (5). In 2010, the system was upgraded to the current version SED 2.0 in order to promote institutional flexibility, integration of cross-curricular subjects into all study plans and curricula known as the Common area (common courses taught throughout all study plans), as well as to update the Institutional Educational Model (IEM) and to promote the educational mission at La Salle.

Evaluation system SED 2.0 includes 3 areas: institutional, educative, and pedagogic. The objective of the institutional area is to evaluate professor's profile from the Lasallian mission (La Salle, 2017). This area applies to all professors of study programs with Recognition of Official Validity of Studies (ROVS). The indicators included in this area are related to personality, community and society. The objective of the educative area is to professor's profile based on abilities of student's graduation profile. Similarly, this area also applies to all professors of study programs with ROVS differentiated by education level (high school, bachelor level, and master level). The indicators included in this area are related to professor's ability of problem solving, social responsibility, ethical judgement, usage of ICT, efficient communication, and information management. The objective of the last are, pedagogic, is to evaluate professor's profile based on his/her capability of learning and teaching. This area applies to all professors of study programs with ROVS differentiated

by education level and type of a course, such as common area, laboratories and workshops, courses of initial phase (first two years), and courses from final phase (last 2 years). (Coordinación de Formación Docente, 2010). The current version of the system consists of 15 questions: institutional area (3 questions), educative (6) and pedagogic (6).

Every student evaluates anonymously all his/her professors from all current courses at the end of each semester. For this purpose, a scale consisting of five options (Never – Almost never – Sometimes – Almost always – Always) is used in the evaluation in each of the three areas. The obtained scores are then transformed to a scale 0-10pts, quantifying the qualitative scale<sup>1</sup>. Students can also add additional comments related to either evaluated subjects, or to administration of studies. The approximate time to finish the evaluation is around 20 minutes. The evaluation is fully available online. Thus, students can make their evaluations at any computer at the university campus, at home, as well as using their mobile phones. The evaluation is opened for 12 days at the end of each semester. During this period, a classroom with 40 computers is reserved at the university especially for the evaluation. What is more, the responsible persons to the evaluation make daily reports to inform the deans of each faculty about the progress in the evaluation (percentage of participation by each study program and group of students).

No more than 10 days after the evaluation, the representatives from DPE deliver the official results to each faculty. The official results include quantitative data (frequencies and evaluation of each professor and each course), qualitative data (students' comments about each professor and about each course), as well as historical data (evaluations and comments since 2010). Every professor has a possibility to revise his/her anonymous evaluation (quantitative and qualitative) on the internet during the first month of the next semester.

## Data

La Salle University México is based in Mexico City, Mexico. Nowadays, the university includes 7 faculties in area of higher education: Mexican School of Architecture, Design and Communication; School of Chemistry; Law School; School of Humanities and Social Sciences; School of Engineering; Mexican School of Medicine; and Business School. Apart of the area of higher education, La Salle University México includes High School and Post graduate studies and Research department. In the academic year 2015/2016, 12,493 students were enrolled in all levels, from which 6,173 students were enrolled in bachelor study programs at all its entities (La Salle, 2016).

In this article, we analyze 517,635 individual evaluations provided by bachelor students between period December 2010 (semester August-December 2010) and December 2016 (semester August-December 2016), i.e. through 13 semesters since the inception of the evaluation system SED 2.0. Table 1 summarizes distributions of the evaluations within the university. These distributions summarize all three areas (institutional, educative and pedagogic) included in SED 2.0. Once the students enter the system, they always terminate the evaluation in all three areas. There is no case, when the evaluation is incomplete. Analyzed 517,635 individual evaluations from SED 2.0 refer to 45,346 professors' evaluations (courses evaluations) throughout the period in question. Table 2 summarizes the descriptive

<sup>1</sup> Each qualitative evaluation is first quantified to a scale 0-4pts, where "Always" corresponds to 4pts. Second, the quantified scale is multiplied by 2.5 to obtain an evaluation on a scale 0-10pts.

statistics of the analyzed sample. In all cases, the results indicate very high average varying from 8.81 at Business School to 9.14 at School of Chemistry in Institutional area, from 8.55 at Business School to 8.99 at School of Medicine in Educative area, and from 8.85 at Business School to 9.1 at School of Medicine in Pedagogic area. The university average is 8.93 in Institutional area (SD of 1.31), 8.72 in Educative area (SD of 1.41), and 8.97 in Pedagogic area (SD of 1.3). Figure 1 summarizes the average evaluations from SED 2.0 for all three areas regarding the whole university and all faculties.

	dec-10	jun-11	dec-11	jun-12	dec-12	jun-13	dec-13	jun-14	dec-14	jun-15	dec-15	jun-16	dec-16
University	36,868	36,383	37,391	36,545	42,190	36,653	43,189	40,200	44,034	41,448	38,500	40,866	43,368
School of Architecture	7,617	7,596	7,410	7,461	9,016	7,880	9,525	9,365	9,848	8,728	8,301	9,239	9,468
School of Chemistry	2,781	2,525	3,364	2,859	3,098	2,307	4,019	3,218	4,067	2,972	3,417	3,003	3,290
Law School	4,138	4,641	5,034	4,898	5,664	5,533	6,440	5,418	6,002	6,303	5,673	5,629	6,221
School of Humanities	2,807	2,717	3,071	2,874	2,905	2,879	2,869	2,964	2,894	2,822	2,343	2,544	2,359
School of Engineering	4,713	3,985	5,180	3,850	5,810	4,495	5,727	4,949	6,041	5,026	5,233	5,195	6,294
School of Medicine	8,022	8,237	5,999	7,525	7,834	6,278	5,991	6,038	6,033	6,439	5,240	5,731	5,279
Business School	6,790	6,682	7,333	7,078	7,863	7,281	8,618	8,248	9,149	9,158	8,293	9,525	10,457

Table 1: Number of evaluations in total, La Salle, December 2010 – December 2016

Regarding Kolmogorov-Smirnov test, the data does respect the normality on confidence interval 95% in all three evaluated areas total, as well as considering each faculty individually as  $p = .000$ . Therefore, Welch-ANOVA is used to analyze differences in the evaluation between faculties at La Salle University México.

The objective of the article is to analyze results from SED 2.0 and get insights into this evaluation. For this purpose, we work with the following three hypotheses:

- $H_0$  – There is no difference in the evaluation with respect to Institutional area and university faculties.
- $H_0$  – There is no difference in the evaluation with respect to Educative area and university faculties.

- $H_0$  – There is no difference in the evaluation with respect to Pedagogic area and university faculties.

The difference in the SED 2.0 evaluation regarding to all three hypotheses is evaluated with respect to mean values and variances.

	N	Institutional		Education		Pedagogic	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
University	45,346	8,932	1,315	8,724	1,419	8,970	1,305
School of Architecture	6,766	8,985	1,186	8,859	1,225	9,007	1,152
School of Chemistry	3,561	9,145	1,230	8,966	1,381	9,056	1,657
Law School	4,413	8,984	1,326	8,805	1,373	8,941	1,327
School of Humanities	3,536	8,944	1,355	8,649	1,519	8,912	1,362
School of Engineering	10,618	8,877	1,354	8,650	1,472	9,040	1,242
School of Medicine	3,919	9,121	1,106	8,992	1,156	9,104	1,103
Business School	12,533	8,810	1,396	8,553	1,510	8,852	1,345

Table 2: Descriptive statistics of SED 2.0 evaluations, La Salle, December 2010 – December 2016 (own calculation)

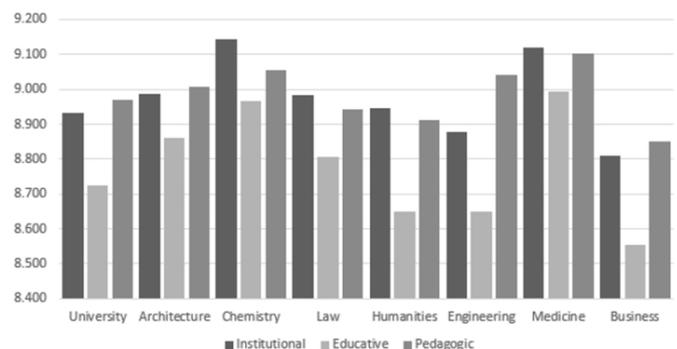


Figure 1: Average evaluation by area and faculty, La Salle, December 2010 – December 2016

**Analysis of variance and Games-Howell test**

The main objective of the article is to analyze results from SED 2.0, i.e. whether there are differences in the evaluation regarding university faculties. For this purpose, we use analysis of variance (ANOVA) with a single factor. We suppose that  $x_1, \dots, x_n, y_1, \dots, y_m$  and  $z_1, \dots, z_m$  are independent samples with mean values  $\mu_x, \mu_y$  and  $\mu_z$ . Further, we suppose that their correspondent variances  $\sigma_x^2, \sigma_y^2$  and  $\sigma_z^2$  are known. We then test following hypothesis

$$H_0 : \mu_x = \mu_y = \mu_z$$

to obtain global significances (Triola, 2012).

If ANOVA indicates statistically significant differences within the analyzed sample, we can apply post hoc test to discover between which groups these differences exist. As the data do not have normal distribution, group sizes are unequal and the variances are not equal, Games-Howell test is applied (Games and Howell, 1976). The Games-Howell test uses formula for Welch's approximate degree of freedom to obtain approximate confidence interval for the difference between two means based on student's *t*-test. We use IBM SPSS Statistics 22 for the analysis.

**Results**

The main objective of the article is to analyze the results of SED 2.0 evaluation, i.e. to understand more what information we can obtain from the evaluation. Nowadays, no deeper analysis about the obtained results is provided. As it was stated in Materials and methods, the Department of Teacher Education (DPE) delivers the official results to each faculty, which are in some cases forwarded to each professor. This transfer of the results is, in most of the cases, done by head of each study program. However, sometimes, this transfer is not done and some professors do not receive his/her evaluations. Similarly, in some case, professors only receive their overall score (usually aggregated evaluation supplemented by separate evaluations from each of the three areas). However, this evaluation does not include either any additional information (e.g. comparative threshold within a study program), or any recommendations and comments from a head of the study program. Thus, the obtain information has lack of added value. In this article, we provide additional analysis of the results to provide better insights into the evaluation.

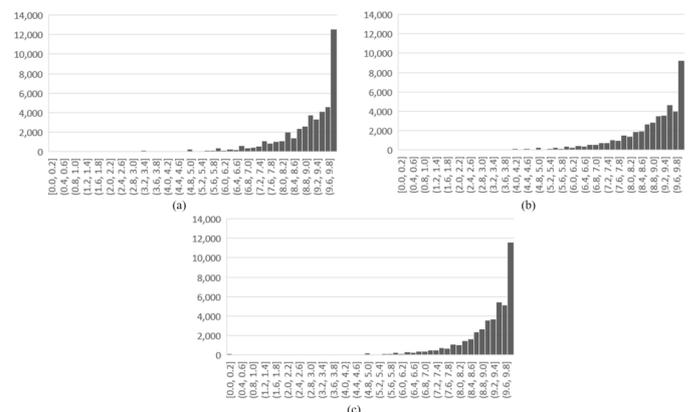
**Distribution of evaluations**

Regarding the scale used in the evaluation (Never – Almost never – Sometimes – Almost always – Always) and its transformation to 0-10pts scale, the obtained averages indicate very positive students' perceptions about professors' teaching quality. Figure 2 shows distributions of evaluations in each area. As we can see, the distribution is significantly skewed to right side (Institutional skew -2.401, Education area -2.15, and Pedagogic area -2.759), i.e. up to the maximum evaluation. When SED 2.0 was created, the expected distribution was rather a normal distribution. The representatives of La Salle University México believed that students would use the whole scale of the evaluation and would more discriminatively express their opinions about the teaching quality.

However, as it is seen in Figure 2, students in most of the cases use highly positive evaluations. Thus, the results of SED 2.0 indicate high teaching quality at La Salle without any problems related to the evaluated areas. The distribution of the evaluations (Figure 2) corresponds to distributions at all its faculties (see Appendix and Figure 3 to Figure 9). The pattern of distributions is skewed to the upper bound of the evaluation (10pts). As a result of this, the most common obtained evaluation in all three areas is the maximal 10pts. In average, 21.038% of evaluations in Institutional area are equal to 10pts (Table 3). Similarly, as the distribution is very similar in Educative and Pedagogic areas, 13.662% of the evaluations in Educative area and 16.689% in Pedagogic area are equal to 10pts (Table 4 and Table 5).

This result can be explained in two possible ways. First, students perceive that teaching at La Salle is of a high quality and, thus, evaluate their professors by the maximal evaluation. However, this might not be the case, as deeper analysis of students' comments included in some evaluations do not support this

idea. In some cases, students refer to various types of problems related to their classes. Rather, second, the maximal evaluation can be explained by students' behavior in evaluation, as in many cases students complete the evaluation as quick as possible without deeper thinking. Moreover, in many cases, if a professor is popular, then receives higher evaluation in all questions. On the other hand, professor without authority (or unpopular) receives lower evaluation. As, nowadays, the evaluation is not optional, and in many cases students are obliged to evaluate their professors, they use pure 10pts in every question. However, this presumption (although it is based on personal professors' and students' experience) must be verified with upcoming research. To support this fact, Table 3, Table 4 and Table 5 provide an information regarding cumulative frequencies in all three areas. For example, in the Institutional area, 92.288% of the evaluations at La Salle are greater than or equal to 7.0, and full of 63.589% of evaluations are greater than or equal to 9.0 (Table 3). This result is similar in the other two areas. We must admit that there might not be anything wrong about these results. Students might feel that it is common to evaluate each professor positively. In this case, evaluation around 7pts and 8pts can be perceived as neutral evaluation, and 6pts as negative evaluation. This students' perception can be connected to grading practice at La Salle University México. At each course, students pass a course when they achieve grade between 6 and 10. If students achieve 5 and lower, then they are obliged to paid and pass extra exam at the end of a semester. Therefore, common higher evaluation in SED 2.0 can be seen as common practice, as students do not consider lower points than 5 or 6.



**Figure 2: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, La Salle, December 2010 – December 2016**

	≥ 1.0	≥ 2.0	≥ 3.0	≥ 4.0	≥ 5.0	≥ 6.0	≥ 7.0	≥ 8.0	≥ 9.0	= 10
University	99.775%	99.667%	99.365%	98.893%	98.196%	95.980%	92.288%	83.694%	63.589%	21.038%
School of Architecture	99.808%	99.719%	99.527%	99.113%	98.478%	97.177%	94.118%	86.580%	64.972%	15.268%
School of Chemistry	99.719%	99.663%	99.495%	99.186%	98.512%	96.883%	94.019%	87.840%	73.069%	27.268%
Law School	99.728%	99.569%	99.230%	98.822%	98.323%	96.102%	92.341%	85.203%	66.327%	21.641%
School of Humanities	99.802%	99.717%	99.463%	98.727%	97.936%	95.532%	91.714%	82.834%	64.791%	23.727%
School of Engineering	99.717%	99.652%	99.331%	98.898%	98.248%	95.592%	91.552%	81.870%	60.586%	25.118%
School of Medicine	99.974%	99.923%	99.592%	99.311%	98.648%	97.474%	95.024%	88.620%	71.192%	11.789%
Business School	99.769%	99.593%	99.218%	98.628%	97.798%	95.021%	90.720%	80.675%	59.012%	20.849%

Table 3: Cumulative distribution of evaluations in Institutional area, La Salle, December 2010 – December 2016

	≥ 1.0	≥ 2.0	≥ 3.0	≥ 4.0	≥ 5.0	≥ 6.0	≥ 7.0	≥ 8.0	≥ 9.0	= 10
University	99.740%	99.581%	99.142%	98.492%	97.279%	94.727%	89.968%	79.076%	55.498%	13.662%
School of Architecture	99.793%	99.719%	99.453%	99.010%	98.256%	96.645%	93.187%	83.314%	58.838%	10.508%
School of Chemistry	99.719%	99.663%	99.495%	99.186%	98.512%	96.883%	94.019%	87.840%	73.069%	27.268%
Law School	99.819%	99.615%	99.298%	98.595%	97.598%	95.083%	90.709%	81.668%	58.441%	15.205%
School of Humanities	99.717%	99.491%	98.869%	98.020%	96.776%	93.241%	88.179%	76.867%	54.864%	14.027%
School of Engineering	99.652%	99.501%	99.077%	98.408%	97.080%	94.330%	88.802%	76.399%	52.439%	16.529%
School of Medicine	99.974%	99.898%	99.617%	99.388%	98.546%	97.193%	94.004%	85.813%	64.506%	5.869%
Business School	99.697%	99.497%	98.899%	98.037%	96.497%	93.385%	87.649%	75.010%	49.406%	12.934%

Table 4: Cumulative distribution of evaluations in Educative area, La Salle, December 2010 – December 2016

	≥ 1.0	≥ 2.0	≥ 3.0	≥ 4.0	≥ 5.0	≥ 6.0	≥ 7.0	≥ 8.0	≥ 9.0	= 10
University	99.687%	99.570%	99.257%	98.754%	98.103%	96.266%	93.009%	85.328%	65.452%	16.689%
School of Architecture	99.867%	99.764%	99.453%	99.143%	98.611%	97.266%	94.561%	87.230%	65.888%	12.001%
School of Chemistry	98.315%	98.175%	97.922%	97.220%	96.883%	95.619%	92.951%	87.616%	74.221%	23.701%
Law School	99.751%	99.660%	99.207%	98.595%	97.915%	96.080%	92.182%	84.478%	64.355%	17.312%
School of Humanities	99.717%	99.548%	99.321%	98.699%	97.851%	95.164%	91.629%	83.937%	64.621%	16.770%
School of Engineering	99.774%	99.689%	99.454%	99.021%	98.437%	96.732%	93.907%	86.297%	67.037%	22.160%
School of Medicine	100.000%	99.898%	99.668%	99.439%	98.673%	97.117%	94.795%	88.951%	70.248%	6.022%
Business School	99.777%	99.633%	99.234%	98.612%	97.854%	95.628%	91.550%	82.390%	60.504%	15.687%

Table 5: Cumulative distribution of evaluations in Pedagogic area, La Salle, December 2010 – December 2016

### Analysis of evaluations in Institutional area

First of all, we analyze whether there is a statistical difference between the analyzed groups, i.e. faculties at La Salle University. The  $F$ -value of 53.423 is statistically significant ( $p = .000$ ) and the test of homogeneity of variances is also statistically significant ( $p = .000$ ). Therefore, the differences between faculties are statistically different, i.e. at least one faculty has statistically different evaluation (Table 6). Further, Welch test is also statistically significant ( $p = .000$ ) as degrees of freedom are quite different (Table 6), which is reflected by differences in standard deviations in the Institutional area (Table 2). As a result, we can reject  $H_0$ , as there is statistically significant difference between faculties regarding the Institutional area. As there is at least one statistically significant difference between faculties, we further apply Games-Howell test to analyze between which faculties the statistically significant differences occur.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	550.326	6	91.721	53.423	.000
Within Groups	77,841.953	45,339	1.717		
Total	78,392.280	45,345			
	Statistic	df1	df2		Sig.
Welch	57.962	6	14991.349	.000	

Table 6: ANOVA analysis of SED 2.0 evaluations in Institutional area, La Salle, December 2010 – December 2016 (own calculation)

For the pairwise analysis of differences, we use the Games-Howell post hoc test, as the data do not meet the homogeneity of variances assumption. Table 7 summarizes results in the Institutional area. In majority of results, there are statistically significant differences between faculties, as the analysis of variance indicated. From the other point of view, there are

no differences between School of Chemistry and School of Medicine (average evaluation is 9.144, 9.121 respectively), which have the highest average evaluations from all faculties (Table 2). Further, there are no differences between School of Architecture, Law School and School of Humanities (8.983, 8.944 and 8.988) and between School of Humanities and School of Engineering (8.994 and 8.877). Last but not least, Business School has the lowest average evaluation in Institutional area (8.810). However, not significantly lower than School of Engineering or School of Humanities.

	School of Chemistry	Law School	School of Humanities	School of Engineering	School of Medicine	Business School
School of Architecture	$p = .000$	$p = 1.000$	$p = .723$	$p = .000$	$p = .000$	$p = .000$
School of Chemistry		$p = .000$	$p = .000$	$p = .000$	$p = .978$	$p = .000$
Law School			$p = .844$	$p = .000$	$p = .000$	$p = .000$
School of Humanities				$p = .143$	$p = .000$	$p = .000$
School of Engineering					$p = .000$	$p = .004$
School of Medicine						$p = .000$

**Table 7: Games-Howell test of SED 2.0 evaluations in Institutional area, La Salle, December 2010 – December 2016 (own calculation), significant at 5% level**

**Analysis of evaluations in Education area**

In the Education area of the evaluation, ANOVA shows similar results as in case of the Institutional area. The  $F$ -value of 91.070 is statistically significant ( $p = .000$ ) and the test of homogeneity of variances is also statistically significant ( $p = .000$ ), as the differences between faculties are statistically different between each other (Table 8). Further, Welch test is statistically significant ( $p = .000$ ) as degrees of freedom are quite different (Table 8). Therefore, we can reject  $H_0$ , as there is statistically significant difference between faculties regarding the Education area. As there is at least one statistically significant difference between faculties, similarly as for the Institutional area, we further apply Games-Howell test to analyze between which faculties the statistically significant differences occur.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,087.641	6	181.274	91.070	.000
Within Groups	90,246.162	45,339	1.990		
Total	91,333.804	45,345			
	Statistic	df1	df2	Sig.	
Welch	99.639	6	14,984.614	.000	

**Table 8: ANOVA analysis of SED 2.0 evaluations in Education area, La Salle, December 2010 – December 2016 (own calculation)**

The post hoc test results in the Education area are summarized in Table 9. Similarly as in the Institutional area, there are statistically significant differences between faculties in majority of cases. There are no differences between School of Chemistry and School of Medicine (8.966 and 8.992), both with the highest average evaluations again (Table 2). Further, there is no difference between School of Humanities and School of Engineering (8.648 and 8.65) and School of Architecture and Law School (8.859 and 8.805).

	School of Chemistry	Law School	School of Humanities	School of Engineering	School of Medicine	Business School
School of Architecture	$p = .002$	$p = .352$	$p = .000$	$p = .000$	$p = .000$	$p = .000$
School of Chemistry		$p = .000$	$p = .000$	$p = .000$	$p = .974$	$p = .000$
Law School			$p = .000$	$p = .000$	$p = .000$	$p = .000$
School of Humanities				$p = 1.000$	$p = .000$	$p = .016$
School of Engineering					$p = .000$	$p = .000$
School of Medicine						$p = .000$

**Table 9: Games-Howell test of SED 2.0 evaluations in Education area, La Salle, December 2010 – December 2016 (own calculation), significant at 5% level**

**Analysis of evaluations in Pedagogic area**

In the Pedagogic area, we get similar results as the  $F$ -value of 34.142 is statistically significant ( $p = .000$ ) and the test of homogeneity of variances is also statistically significant ( $p = .000$ ), as the differences between faculties are statistically different (Table 10). Further, Welch test is statistically significant ( $p = .000$ ) as degrees of freedom are quite different (Table 8). Therefore, we can reject  $H_0$ , as there is statistically significant difference between faculties regarding the Pedagogic area. Thus, post hoc test can be applied to analyze between which faculties the statistically significant differences can be observed.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	347.462	6	57.910	34.142	.000
Within Groups	76,899.394	45,339	1.696		
Total	77,246.856	45,345			
	Statistic	df1	df2	Sig.	
Welch	35.166	6	14,744.893	.000	

**Table 10: ANOVA analysis of SED 2.0 evaluations in Pedagogic area, La Salle, December 2010 – December 2016 (own calculation)**

The following Table 11 summarizes the Games-Howell test results in the Pedagogic area. Although the results indicate statistically significant differences between most of the faculties, we can observe less significant results than in the previous two areas. There are no differences between School of Architecture, School of Chemistry, Law School and School of Engineering, as well as between School of Chemistry and School of Medicine, and School of Humanities and Business School.

	School of Chemistry	Law School	School of Humanities	School of Engineering	School of Medicine	Business School
School of Architecture	$p = .707$	$p = .096$	$p = .007$	$p = .584$	$p = .000$	$p = .000$
School of Chemistry		$p = .014$	$p = .001$	$p = .998$	$p = .769$	$p = .000$
Law School			$p = .960$	$p = .001$	$p = .000$	$p = .002$
School of Humanities				$p = .000$	$p = .000$	$p = .241$
School of Engineering					$p = .042$	$p = .000$
School of Medicine						$p = .000$

**Table 11: Games-Howell test of SED 2.0 evaluations in Pedagogic area, La Salle, December 2010 – December 2016 (own calculation), significant at 5% level**

## Discussion

The analysis of the obtained evaluations between December 2010 and December 2016 from SED 2.0 indicates positive perception of teaching quality at La Salle University México. The results show significant similarities in all three areas of SED 2.0, as well as between all faculties. We can accept the results with a conclusion of well-established evaluation system and positive teaching quality. However, there are several tasks that should be taken into consideration for future development of both SED 2.0 and teaching at La Salle. For example, how to increase discrimination ability in the evaluations and receive more accurate results.

First, we should consider reevaluating the current structure of SED 2.0, i.e. whether the evaluation structure in Institutional, Educative and Pedagogic areas is correct regarding students' expectations. For example, whether questions included in the system correspond with current students' needs, as it is of a high importance that an evaluation system complies with students' interest. As Brand Barajas (2014) pointed out, any evaluation must satisfy perfectly the objectives of the evaluation, as well as the necessities and interests of those who realize the evaluation (in our case the interest of students). Thus, the internal motivation and interest are the key feature to provide an evaluation of a high quality. According to McAuley et al. (2017), the top three most important motivators for students to finish an evaluation are: 1) to make course better, 2) earn bonus points, and 3) improve professor's teaching. On the other hand, the top three barriers to finish an evaluation are: 1) evaluate multiple professors, 2) complete several evaluations at the same time, and 3) complete lengthy evaluations. Regarding these findings, La Salle should consider changes in SED 2.0 structure that would deal with these motivators and barriers.

Moreover, besides structural changes in question, we can also consider changes in the evaluation scale in each question. The current version of SED 2.0 uses five-point scale (Never – Almost never – Sometimes – Almost always - Always) in each question. Obtained evaluation is then transformed to 0-10 points scale using a specific algorithm. As the discrimination ability is not satisfactory, a wider scale would provide results with better distribution within the evaluation scale. With a wider scale students may feel more freedom to express their opinions in the evaluation. However, as Flégl et al. (2017) investigated, wider scale does not necessarily need to lead to better discrimination within an evaluation. Students' behavior during the evaluation influences more the results than type of a scale. Thus, any changes in the evaluation scale must be carefully considered whether it would have a positive impact. Moreover, we still must keep in mind that any change must go along with the objective of the evaluation itself, as well as along with students' opinions. From this reason, we have prepared a questionnaire to investigate students' opinions about SED 2.0 and usefulness of the evaluation. This questionnaire was disseminated at all faculties in February, March and April 2017. According to our plan, the obtained results will be analyzed within the following months. If we gain valuable information of students' opinion about the system of professors' evaluation, then we can modify the current structure of the system and obtain more precise evaluation of each course later on. The more precise evaluations can be later used for improving teaching quality at La Salle. Consequently, precise results can serve to improvements of a course structure, as well as to professional development of a professor. As Marsh and Roche (1993) and Santibañez (2006) emphasize the improvement of teaching quality is directly connected to educative quality at an institutional level. Thus,

positive synergy in education process can be obtained at all parts of La Salle University México.

Second, connection between completion of the evaluation at the end of each semester and students' grading can also be considered. Nowadays, SED 2.0 is not connected to students' grading at all. Nevertheless, students are "obliged" to make the evaluation. So, a discussion can arise whether make the completion optional, free on each student's decision, and, somehow, connected to their grading. This change might have a positive effect, as well as negative one. Make the evaluation optional might directly lead to a drop in the total number of completed evaluations. Last year, SED 2.0 recorded 43,368 individual evaluations (Table 1) within a population of 6,173 students enrolled in bachelor study programs (four-semester long average of 41,046 individual evaluations). Potential drop can be significant in the following period after the change is made (hardly predictable). However, after the initial decrease, this drop can consolidate and the total number of completed evaluations can begin rising again. What is more, we might obtain better distributed evaluations as only those students interested in the evaluation would express their opinions about teaching quality.

In addition, to prevent the drop in the evaluations, we can consider connecting the optional evaluation to on-line publication of students' semester results (incentive for students). In this case, for example, only those students that complete the evaluation can be able to see their final grades on-line, i.e. accessible from outside of the university campus. On the other those, who do not complete the evaluation must come to university to see their final grades. This might be another stimulus for a completion of the evaluation. We can see this as an incentive, which will not improve grading obtained in courses during a semester. This possibility must be again carefully analyzed, as students can fill in the evaluation rapidly without deeper thinking just to have the option to see their results on-line.

Third, we should consider improving reporting of the results from SED 2.0. In the current system, there is no automatic reporting system, which sends professor's evaluation directly to a professor. The results are first proceeded by authorities at Department of Teacher Education, and after that sent to all heads of study programs. Unfortunately, not all heads of the study programs forward the evaluation to all professors. What is more, the information that is sent only includes overall evaluation (with anonymous students' comments if any) without any comparative threshold. Thus, this information lacks additional value that would serve as a base for potential teaching improvements. As Bolívar (2008) pointed out, if we do not have a synergy between different parts at university, then we cannot expect improvements in teaching quality. For example, if professors do not have valuable feedback regarding their evaluation, then there cannot be any improvement in teaching quality. Therefore, we should consider involving statistical analysis onto SED 2.0 to provide comparative evaluation thresholds. In this case, each head of a study program and each professor would be able to see how the evaluation ranks, for example, at a faculty. Having in hands comparative analysis, the representatives at university would offer specific feedback to professors, as well as additional training if needed.

Although we consider changing SED 2.0 in a way described above, we must always consider few specifics that influence professors' evaluations. The result of an evaluation depends mostly on students' interest about a course, the ease of the course and on gender of the evaluated professor (Leung, Jiang and Busser, 2013). The higher the interest is, the higher the evaluation is. Therefore, it is up to a professor to make a course

interesting for its students. Furthermore, as students evaluate their classes before the end of a semester, we must consider that expected course grade has also a large effect on the evaluation (Langbein, 2008). Students like to be satisfied, and they like higher grades. However, the ease of a course can have both a positive impact, as well as negative impact on an evaluation (Felton et al., 2008, Marsh and Roche, 2000). The evaluation depends on cultural aspects of a country and cultural aspects of a university.

Although the effect of ease of a course has an impact on the evaluation, this effect can be diminished by fairness grading procedure (Wendorf and Alexander, 2005). Thus, if students feel grading procedure is fair and sufficiently clear, then this fairness has bigger impact on professors' evaluation than the ease of class. Therefore, a perception about better evaluation in case of "easier" courses is not the unique influential factor in professors' evaluation. For example, gender of a professor has also direct impact on an evaluation. In general, male professors usually get higher evaluation (Basow, 2000). Similarly, the attractiveness of a professor has positive effect on the evaluation (Felton et al., 2008, Silva et al., 2008). Considering all of these reasons, it is of a high importance to analyze all these effects on the professors' evaluation before initiating structural changes in the evaluation system.

## Conclusion

The article deals with introductory analysis of the results of professors' evaluation at La Salle University México. As no additional analysis of the results had been created since the inception of the evaluation system (2010), this article provides the initial insights. At La Salle and its faculties, the evaluation is highly skewed towards the maximal evaluation of 10 points. In general, students evaluate their professors (courses) positively, which may indicate no problems regarding teaching quality. Approximately, 80% of the evaluations are greater than or equal to 8.0. However, students' additional comments included in the evaluation indicate several problems. This highly positive evaluation can have several reasons, such as common qualification scale used at La Salle (to pass a course a grade between 6 and 10 must be reached). Thus, students might not, unconsciously, be willing to evaluate professors at the whole scale.

Further, the analysis indicates statistically significant differences in the evaluation between faculties and, thus, research hypotheses were rejected. The highest average evaluation can be observed at School of Chemistry and Mexican School of Medicine. Nevertheless, statistically significant differences exist between more faculties. This introductory analysis is based on only one factor. To be able to explain reasons of these differences, further analysis focusing on effects that directly influence the evaluation must be done. The future research will focus whether factors such as gender, seniority, attractiveness, etc. have impact on the evaluation. Similarly, the future analysis will focus on the structural changes within the evaluation system towards new version SED 3.0.

## Acknowledgements

This article was supported by Universidad La Salle México under Grant "Análisis de rendimiento y eficiencia en el área educativa institucional", number EDU-09/17.

## References

- Basow, S.A. (2000) 'Best and Worst Professors: Gender Patterns in Students' Choices', *Sex Roles*, vol. 43, no. 5, pp. 407-417. <http://dx.doi.org/10.1023/A:1026655528055>
- Becker, W.E., Watts, M. (1999) 'How departments of economics should evaluate teaching', *American Economic Review (Papers and Proceedings)*, vol. 89, no. 2, pp. 344-349. <http://dx.doi.org/10.1257/aer.89.2.344>
- Bolívar, A. (2008) 'Evaluación de la práctica docente. Una revisión desde España', *Revista Iberoamericana de Evaluación Educativa*, vol. 1, no. 2, pp. 57-74.
- Braga, M., Paccagnella, M., Pellizzari, M. (2014) 'Evaluating students' evaluations of professors', *Economics of Education Review*, vol. 41, pp. 71-88. <http://dx.doi.org/10.1016/j.econedurev.2014.04.002>
- Brand Barajas, J. (2014) *La evaluación de los coordinadores de programas académicos: una propuesta de indicadores. La evaluación educativa desde diversos espacios académicos*, Universidad La Salle México, DeLaSalle ediciones, 1ª edición, México.
- Coordinación de Formación Docente (2010) 'Documento interno de trabajo', Universidad La Salle México, México.
- Felton, J., Koper, P.T., Mitchell, J., Stinson, M. (2008) 'Attractiveness, easiness and other issues: student evaluations of professors on Ratemyprofessors.com', *Assessment & Evaluation in Higher Education*, vol. 33, no. 1, pp. 45-61. <http://dx.doi.org/10.1080/02602930601122803>
- Flegl, M., Roa Rivas, K.X., Vergara Abascal Sherwell, G., Monroy Jiménez, M.A. (2017) 'On alternative methodology for increasing discrimination ability in system of professors' evaluation', *Proceedings of Efficiency and Responsibility in Education (ERIE) 2017*, pp. 72-79.
- Games, P.A., Howell, J.F. (1976) 'Pair wise multiple comparison procedures with unequal n's and/or variances', *Journal of Educational Statistics*, vol 1., no. 2, pp. 113-125. <http://dx.doi.org/10.2307/1164979>
- Hein, N., Kroenke, A., Rodrigues Júnior, M.M. (2015) 'Professor Assessment Using Multi-Criteria Decision Analysis', *Procedia Computer Science*, vol. 55, pp. 539:548. <http://dx.doi.org/10.1016/j.procs.2015.07.034>
- La Salle (2017) *Modelo Educativo - Universidad La Salle, Ciudad de México*, [Online], Available: <http://www.lasalle.mx/somos-la-salle/modelo-educativo/> [11 May 2017].
- La Salle (2016) *Quinto INFORME del Rector 2015-2016*, Universidad La Salle México, México.
- Langbein, L. (2008) 'Management by results: Student evaluation of faculty teaching and the mis-measurement of performance', *Economics of Education Review*, vol. 27, no. 4, pp. 417-428. <http://doi.org/10.1016/j.econedurev.2006.12.003>
- Leung, X.Y., Jiang, L., Busser, J. (2013) 'Online student evaluations of hospitality professors: A cross-cultural comparison', *Journal of Hospitality, Leisure, Sport & Tourism Education*, vol. 12, no. 1, pp. 36-46. <http://dx.doi.org/10.1016/j.jhlste.2012.10.001>
- Marsh, H.W., Roche, L. (1993) 'The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness', *American Educational Research Journal*, vol. 30, no. 1, pp. 217-251.
- Marsh, H.W., Roche, L. (2000) 'Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders?', *Journal of Educational Psychology*, vol. 92, no. 1, pp. 202-228. <http://dx.doi.org/10.1037//0022-0663.92.1.202>
- McAuley, J.W., Backo, J.L., Sobota, K.F., Metzger, A.H.,

- Ulbrich, T. (2017). 'Identifying motivators and barriers to student completion of instructor evaluations: A multi-faceted, collaborative approach from four colleges of pharmacy', *Currents in Pharmacy Teaching and Learning*, vol. 9, no. 1, pp. 20-27. <http://dx.doi.org/10.1016/j.cptl.2016.08.029>
- Santibañez, L. (2006) 'Why we should care if teachers get A's: Teacher test scores and students achievement in Mexico', *Economics of Education Review*, vol. 25, no. 5, pp. 510-520. <http://dx.doi.org/10.1016/j.econedurev.2005.08.001>
- Silva, K.M., Silva, J.F., Quinn, M.A., Draper, J.N., Cover, K.R., Munoff, A.A. (2008) 'Rate My Professor: Online Evaluations of Psychology Instructors', *Teaching of Psychology*, vol. 35, no. 2, pp. 71-80. <http://dx.doi.org/10.1080/00986280801978434>
- Triola, M.F. (2012) *Elementary Statistics*, 12th edition, Pearson.
- Wendorf, C.A., Alexander, S. (2005) 'The influence of individual- and class-level fairness-related perceptions on student satisfaction', *Contemporary Educational Psychology*, vol. 30, no. 2, pp. 190-206. <http://dx.doi.org/10.1016/j.cedpsych.2004.07.003>

Appendix

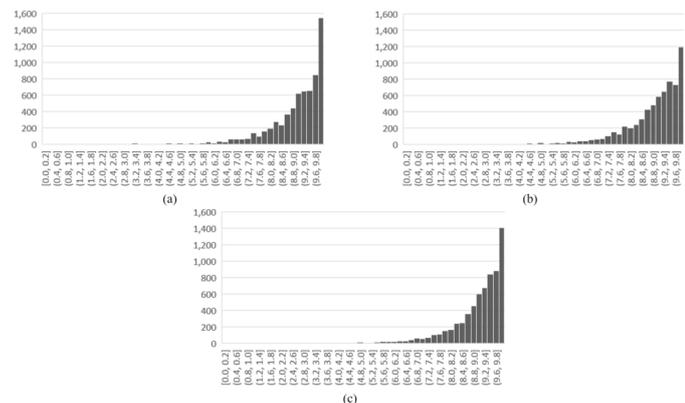


Figure 3: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, School of Architecture, December 2010 – December 2016

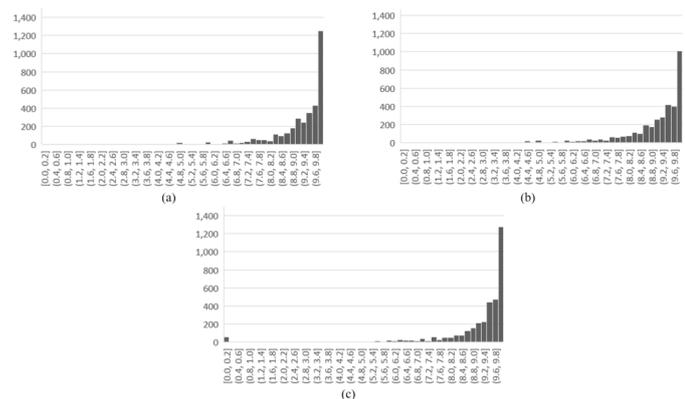


Figure 4: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, School of Chemistry, December 2010 – December 2016

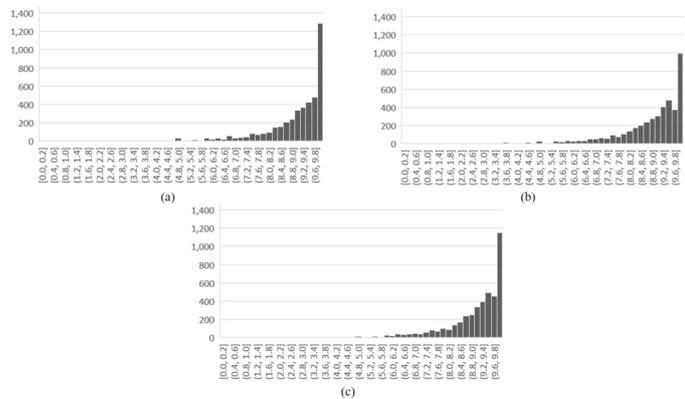


Figure 5: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, Law School, December 2010 – December 2016

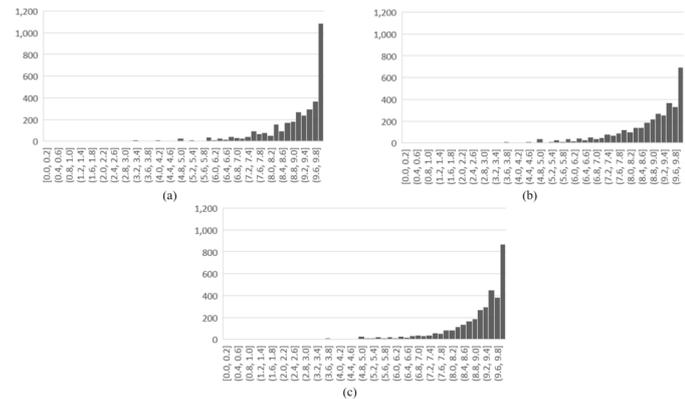


Figure 6: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, School of Humanities, December 2010 – December 2016

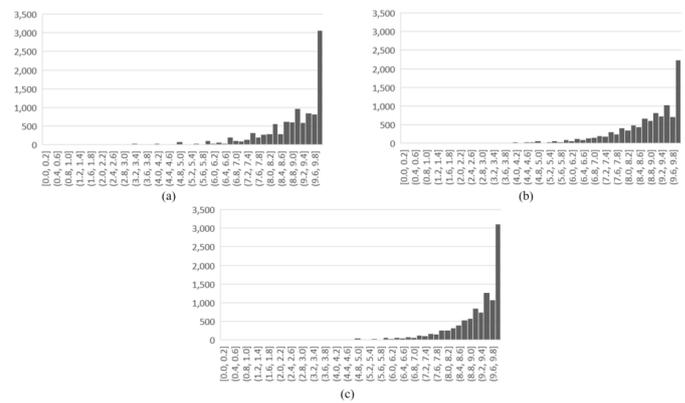


Figure 7: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, School of Engineering, December 2010 – December 2016

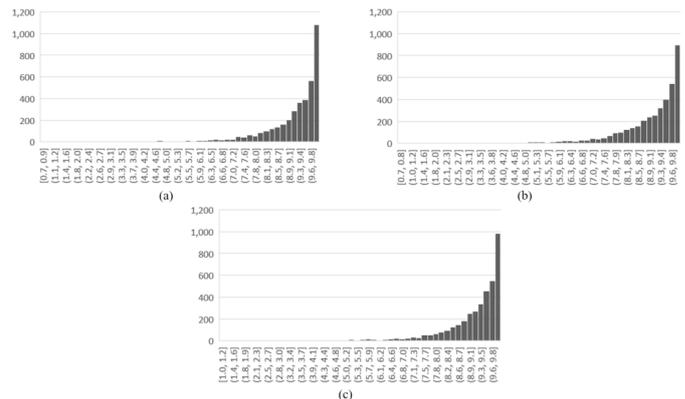


Figure 8: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, Mexican School of Medicine, December 2010 – December 2016

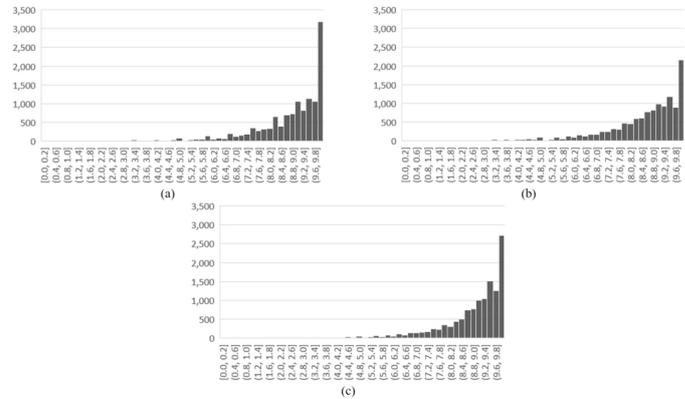


Figure 9: Distribution of evaluations in a) Institutional, b) Educative, and c) Pedagogic area, Business School, December 2010 – December 2016