

Investigating Student Evaluation of Teachers by Using Latent Class Analysis: A Case Study at a Tertiary Level

Evrım Üstünlüoğlu^{1,*} & Duygu Güngör-Culha²

¹School of Foreign Languages, İzmir University of Economics, Sakarya Street No:156 35330 Balçova İzmir, Turkey

²Department of Psychology, İzmir University, Gürsel Aksel No:14 İzmir, Turkey

*Corresponding author: Evrim Üstünlüoğlu, School of Foreign Languages, İzmir University of Economics, Sakarya Street No:156 35330 Balçova İzmir, Turkey. E-mail: evrim.ustunluoglu@ieu.edu.tr

Received: May 16, 2012 Accepted: June 19, 2012 Published: September 22, 2012

doi:10.5296/ije.v4i3.1811 URL: <http://dx.doi.org/10.5296/ije.v4i3.1811>

Abstract

Student evaluations of teaching and teacher (SET) have become the focus of extensive data collection, due to high levels of competition in education. Yet, analysis of the data collected from students has been relatively neglected in favor of evaluation. Furthermore, heterogeneity of student population with different academic background may require advanced statistical techniques, such as Latent Class Analysis (LCA), used in the context of Latent Variable Models, to disclose latent classes or structures by the manifest variables. The purpose of the study is to identify distinct groups of students based on their SET ratings and use the LCA method to discover whether there is a discrepancy in the identified classes in terms of level of success and gender. The study also aims to present a descriptive examination regarding the students who evaluated the instructors and which classes they belonged to. The following three conclusions have been drawn from this research; that different typological structures of student exist in the institution, that there are differences in the identified classes in terms of gender, and that, regardless of whether they were successful or not, students were generally positive about teachers.

Keywords: *Student evaluation; Teaching; Latent Class Analysis*

1. Introduction

In spite of disagreement over consequences regarding their positive and negative impact on teaching and learning, Student Evaluations of Teaching and Teachers (SET) have been widely used, particularly in tertiary institutions for years (Jones, 1989; Ory & Ryan, 2001). Thus, the feedback from students has become common and is regarded as valuable for many institutions. Studies have mainly focused on the factors which might have an impact on validity and reliability of student evaluations, defining three major categories: instructor-level, subject-level and student-level (Pozo-Munoz, Rebollosa-Pacheco & Fernandez-Ramirez, 2000).

Instructor- level factors are summarized as instructors' use of class time, availability outside class time, how well they assess student learning or understanding, concern for students' welfare and performance, the extent to which they emphasize analytical or critical skills, preparedness, and tolerance of alternative viewpoints in class (Pozo-Munoz et al., 2000). According to Boex (2000), presentation and organizational skills, clarity of expression, how the instructor uses grading and assignments, intellectual capabilities, the ability to interact well with students, and the ability to motivate students are also important instructor level factors. A number of studies also highlight the significance of instructor personality for SET, including expertise, ability to motivate, management of student behavior, level of excitement, interpersonal skills, showing a caring nature, being systematic, and showing respect for students (Brown & Atkins, 1993; Lowman & Mathie, 1993; Patrick & Smart, 1998). In addition to these, instructor's reputation was also found to influence ratings, unlike title and position (Boex 2000; Jacobs 2002; Murray, Rushton & Paunonen, 1990; Shevlin et al., 2000).

Subject-level factors in evaluations of students include the time of the lesson, whether it is elective or a prerequisite, the level, perceived difficulty, and the size of the class (Neumann, 2000). Although time of the day or year do not seem to have a great effect on evaluation, rating surveys given immediately after a final exam have been found to have less validity because of anxiety and fatigue. Regarding class size, it has been found that students in larger classes give lower ratings. In their extensive study, Davies et al. (2006) found that the class size did not have a significant effect on student rating in first and second year subjects, but had a negative effect in later year classes. Whether subject itself has an effect on evaluation or not has also been a focus of studies. It has been found that Humanities and Art type subjects receive higher ratings than Mathematics-type courses, as students tend to feel incompetent in quantitative skills (Braskamp & Ory, 1994; Cashin, 1990; Neumann, 2000).

Student-level factors include student biases, reasons of taking the course, the effort student expend in the subject, age, ethnicity, gender, and students' grade expectations. Studies on gender have yielded inconclusive results, as some support potential biases, while others indicate that gender has no effect on evaluation (Basow, 2000; Cashin, 1990; Feldman, 1993; Wolfer & Johnson, 2003). However, Mason, Steagall and Fabritus (1995) claim that females are more likely to give positive ratings of teacher effectiveness. Regarding age, no significant effect has been found on evaluation, however, Worthington (2002) claims that

students over 30 are more likely to give lower ratings. Among all factors, student expectations seem to be the center of several studies. Focusing on grade expectations in his study, Marsh (1987) grounds his opinion on correlation between grades and student evaluations in three hypotheses 1- the leniency hypothesis, which is related with teacher's leniency in grading, 2- the validity hypothesis, which is related with the amount of knowledge students have gained and the favoritism that they show, by giving high rates, and 3- the prior characteristic hypothesis, which is related with particular student or course factors such as motivation or class size. Other factors arising from student, such as cultural background, thinking style, learning style, high grade expectation, and nationality have also been found to be positively correlated with instructor ratings, although they are not directly related to the instructor (Boex, 2000; Germain & Scandura, 2005; Zhang, 2004; Worthington, 2002). Underlying the importance of nationality, Worthington (2002) claims that students from a non-English speaking background expect higher grades and tend to give higher ratings (Millea & Grimes, 2002). Personality, mentioned as instructor level factor above, is also regarded as student level factor, since students tend to give higher rates to the teachers who are kind, funny, enthusiastic and entertaining (Feldman, 1993; Wilson, 1998).

In addition to the instructor, student and subject level factors mentioned above, the method of administration applied in SET is also important for reliability purposes. A great many studies shows that anonymity has an effect on students' evaluations of teachers, and that non-anonymous methods result in higher ratings than anonymous ones (Fries & McNinch, 2003). Davies et al. (2010) indicate that the purpose, the content and the type of surveys are important factors, and these vary from institution to institution, which might cause different interpretations. The purpose of evaluation makes a difference to students' ratings, some universities conduct surveys to collect qualitative data specifically on lecturer and subject characteristics, others, on student learning. Students have been found to give lower ratings when the purpose is given as tenure, but higher ratings for when given as the improvement of teaching (Worthington, 2002).

As indicated in the studies above, student evaluations of teaching and teacher have become the focus of extensive data collection, due to high levels of competition in education. Yet, analysis of the data collected from students has been relatively neglected in favor of evaluation. When related studies are carefully examined, it can be seen that the analysis of data usually focuses on frequency distributions and correlations, although the purpose is to identify the latent variables in this type of evaluation. However, heterogeneity of student population with different academic background may require advanced statistical techniques, such as Latent Class Analysis (LCA), used in the context of Latent Variable Models, to disclose latent classes or structures by the manifest variables. In other word, LCA is a statistical method for identifying and analyzing unmeasured class membership among subjects using categorical and/or continuous observed variables (McCutcheon, 2002), and thus, provides a way of grouping individuals on the basis of shared characteristics that distinguish members of one group from those of another (Collins & Lanza, 2010). Different from other latent variable models, LCA is a person-centered method and the focus is on relationships among individuals (Muthen & Muthen, 2000). Considering that student

evaluations of teaching and teacher could include a number of latent structures, the purpose of using LCA in this study is based on three reasons: to classify students into discrete groups based on observed variables, to reveal underlying patterns to develop a deeper understanding of the data, and to identify whether there are different student typologies based on SET ratings. It is also important to note here that LCA has two parameters which help interpreting the classes and their nature in this study. One is latent class probabilities, which give the number of classes and the percentage of students which fall into each; and the other is conditional probabilities, which help to interpret the nature of classes, such as whether they are perceived negatively or positively. A survey of studies conducted on SETs reveal that an analysis of SETs using LCA is rather uncommon; therefore, this study may be valuable in creating a different perspective, resulting in potentially more profound inferences.

1.1 Purpose of the Study

The purpose of the study is to identify distinct groups of students based on their SET ratings and use the LCA method to discover whether there is a discrepancy in the identified classes in terms of level of success and gender. The study also aims to present a descriptive examination regarding the students who evaluated the instructors and which classes they belonged to.

2. Method

2.1 Participants

923 English Medium University, Preparatory Program students at tertiary level participated in this study. Of these, 460 were male and 463 were female with ages ranging from 18 to 22, and English levels varying from Pre-Intermediate to Upper Intermediate.

83 English instructors -18 male and 65 female- were evaluated by the participants. Of these, 62 were Turkish and 21 were native speakers of English. Instructors' teaching experience varied from three to 20 years.

2.2 Data Collection

The university where the study was conducted has had a student evaluation of teaching and instructor system in place for eight years. The purpose of student evaluation is to support the improvement of teaching effectiveness and quality. The students are informed about the purpose and procedures of evaluation during orientation sessions given in the first week of every academic year. The data in this system are collected anonymously through an electronic questionnaire at the end of each academic term: fall and spring. The Likert-type questionnaire, which was developed for this purpose by a committee of experts, has been piloted and revised; however it is updated when deemed necessary based on feedback from students and teachers. The questionnaire includes 10 items regarding instructor's in-classroom teaching and management roles.

The data in this study were collected at the end of the Spring Term before their final exam. The questionnaire originally has five rating scale, from strongly agree to strongly

disagree. However, although LCA method analyzes five category questions, in this study category number was collapsed to three –strongly agree, undecided, strongly disagree- in order to allow more accurate interpretation. Out of 10 items in the questionnaire, seven (Note 1) was selected and represented in Table 1.

The data regarding students' success level were collected from the English exam given at the end of Spring semester. The exams which are regularly given at the end of Fall and Spring semester are prepared by the Testing Office, based on the specifications identified in line with the objectives based on student needs before the Academic Year. The Testing Office regularly works on test reliability and validity, carefully following the procedure for standardization. The pass grade of the exam is 65 and in this study, 313 passing students were considered successful, whereas 610 who failed were considered unsuccessful.

2.3 Data Analysis

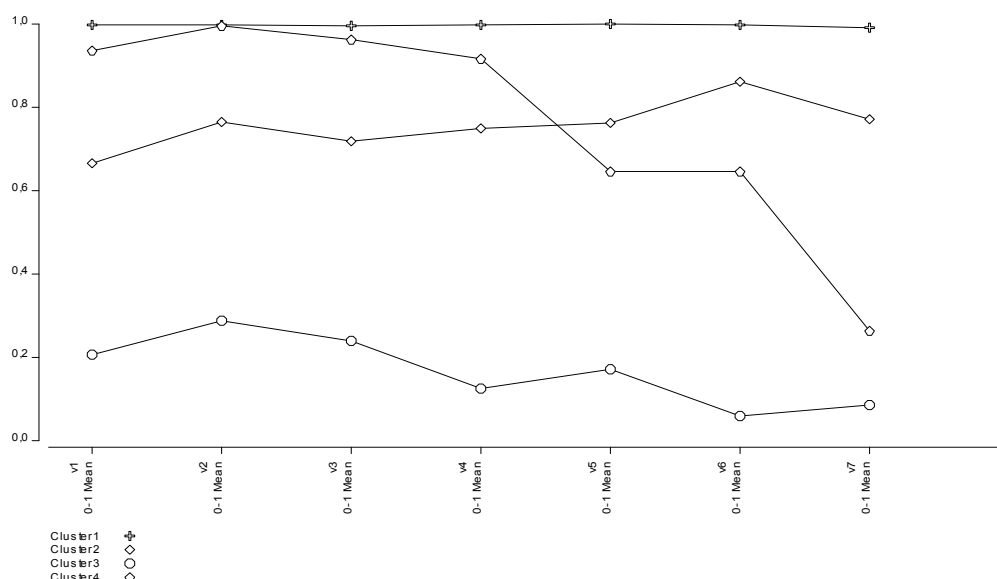
In order to analyze the data, first, student evaluation data were submitted to series of latent class models using a comprehensive modeling program, Latent GOLD 4.5, developed by Vermunt and Magidson (2005). The five models including 1-5 classes were analyzed by using seven indicators. Both Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC) were used for model selection. In the study conducted by Nylund, Asparouhov and Muthen (2007), the BIC was found to perform better than other information criteria and likelihood ratio tests in identifying the appropriate number of latent classes. Although different formulas are used in the calculation of the criteria by computer, it is generally accepted that better fitting models have lower information criteria values. After model selection, the properties of the classes found were discussed. Third, χ^2 and ANOVA statistics were performed to test for differences in gender and success of students between clusters. Finally, frequencies were calculated and presented.

3. Findings and Discussion

The first purpose of the study was to identify distinct groups of students based on their evaluation ratings. The results of the analysis indicated four distinct groups with the lowest BIC and AIC, which was found as the best fitting model (respectively 4913.13 and 4729.68 see Table 1). The entropy score also indicated that 90% of students were correctly classified by using 4-class model. Parameter estimates of four class model – latent class probabilities and conditional probabilities- are presented in Table 2 and also visualized in Figure 1.

Table 1: Model Fit Estimates for 1 to 5 Clusters

Number of classes	LL	BIC(LL)	AIC(LL)	Npar	L ²	df	p-value	Class.Er.
1-Cluster	-3730,41	7556,41	7488,825	14	3545,09	909	1,10E-306	0
2-Cluster	-2614,15	5378,49	5272,291	22	1312,56	901	7,00E-18	0,0097
3-Cluster	-2367,63	4940,09	4795,264	30	819,533	893	0,96	0,0184
4-Cluster	-2326,84	4913,13	4729,683	38	737,952	885	1	0,0279
5-Cluster	-2303,63	4921,33	4699,267	46	691,535	877	1	0,033



V1-V7 can be seen from table.2.

Figure 1: 4-Class Model

Table 2: Parameter Estimates of 4-Class Model(re-expressed as conditional probabilities)

		Cluster1	Cluster2	Cluster3	Cluster4
Overall	Latent Class Probabilities	0,7303	0,1116	0,0865	0,0716
Indicators					
<i>V1-Knows his/her topic</i>					
don't agree		0	0,2132	0,7341	0,0065
Undecided		0,0088	0,3655	0,1845	0,1773
Agree		0,9912	0,4213	0,0814	0,8162
<i>V2-Uses tools and materials in a timely and appropriate manner</i>					
don't agree		0	0,1267	0,6381	0
Undecided		0,0088	0,3299	0,2252	0,0168
Agree		0,9912	0,5434	0,1366	0,9832
<i>V3-Corrects wrong and incomplete answers effectively</i>					
don't agree		0	0,148	0,6852	0,0009
Undecided		0,0136	0,4011	0,2261	0,1128
Agree		0,9864	0,4509	0,0887	0,8864
<i>V4-Concludes and summarizes the lesson clearly</i>					
don't agree		0	0,0776	0,8248	0,0024
Undecided		0,0084	0,5201	0,1551	0,248
Agree		0,9916	0,4023	0,0201	0,7495
<i>V5-Ensures active participation of the students</i>					
don't agree		0	0,1062	0,7727	0,2175
Undecided		0,0033	0,3952	0,1735	0,4118
Agree		0,9967	0,4987	0,0538	0,3707
<i>V6-Stimulates the interest to the lesson</i>					
don't agree		0	0,0452	0,9211	0,2391
Undecided		0,0089	0,2798	0,0629	0,3501
Agree		0,9911	0,675	0,0159	0,4107
<i>V7-Praises student effort</i>					
don't agree		0	0,093	0,8812	0,6518
Undecided		0,0277	0,4118	0,0997	0,2563
Agree		0,9723	0,4952	0,0192	0,0919

The results, based on parameters, indicated that students in Cluster 1 (73%) had a high likelihood of agreeing with the items from one to seven in the questionnaire. This means that these students consider that instructors know their topic, use tools and materials in a timely and appropriate manner, correct wrong and incomplete answers effectively, conclude and summarize the lesson clearly, ensure active participants of the students, stimulate the interest to the lesson, and praise student effort. This group can be labeled as *the positive/positive group*. The students in Cluster 2 (11%) tended to have no definite views on these items, therefore, can be labeled as the *undecided/undecided group*. The results indicated that students in Cluster 3 (8%) disagree with all items indicated in the questionnaire, and can be labeled as the *negative/negative group*. The results also indicated that the student in Cluster 4 (7%) generally agreed with items regarding instructor's in-classroom teaching activities, such as knowing his/her topic, using tools and materials in a timely and appropriate manner, correcting wrong and incomplete answers effectively, and concluding and summarizing lesson clearly. However, this group seemed to be "undecided" about instructor's ensuring active participation of students, and they also seemed "not agree" with instructor's praising student effort. An analysis of these figures reveals satisfaction with overall teaching ability, but suggests the classes were focused on the teacher rather than students. This group can be labeled as *unrewarded* since they perceived their instructors did not adequately praise their efforts (see Table 2).

The second purpose of the study was to find out whether there was any variation in the identified classes in terms of gender and success. The results indicated no significant difference among the classes in terms of their success, $F(3,911)=0.4$ $p>0.05$. It is important to note here that of the 923 students who evaluated their instructors, 313 were successful, while 610 failed the exam. Although many students failed in the exam, the majority of these gave positive evaluation, as the successful students did. Regarding gender, a significant difference was found among the groups, $\chi^2(3, N=923)=21,32$, $p<0.001$. As seen at Table 3, Cluster 4 had a significantly greater proportion of females. It is interesting to note that this female dominated group had positive perceptions of teaching ability, but were affected by the lack of praise given (see Table 3).

Table 3: Differences between Classes in terms of gender and success

	Cluster 1 (n=680)	Cluster 2 (n=97)	Cluster 3 (n=81)	Cluster 4 (n=65)	F	dfs	$\chi^2(3)$
Student Success ^a	53,28	53,84	52,87	54,76			
M (SD)	(12,20)	(11,74)	(12,00)	(9,05)	0,4	3, 911	
Student Sex ^b N(%)							
Female	343 (74,1%)	39 (8,4%)	33 (7,1%)	48 (10,4%)			
Male	337 (73,3%)	58 (12,6%)	48 (10,4%)	17 (3,7%)			21,327***

^a Exam scores of students were taken as student success.

^b Distribution of students to the clusters according to gender

* $p < ,05$ ** $p < ,01$ *** $p < ,001$

Table 4: Frequencies (%) of students based on their classes

Cluster	1	2	3	4	Cluster	1	2	3	4
Instructor1	79,17%	10,42%	6,25%	4,17%	Instructor44	48,10%	15,19%	31,65%	5,06%
Instructor2	50,00%		50,00%		Instructor45	42,86%	28,57%	28,57%	
Instructor3	100,00%				Instructor46	33,33%		66,67%	
Instructor4	50,00%	50,00%			Instructor47	90,00%	10,00%		
Instructor5	75,00%	12,50%		12,50%	Instructor48	93,10%	6,90%		
Instructor6		100,00%			Instructor49	50,00%	50,00%		
Instructor7	83,33%	16,67%			Instructor50	63,64%		9,09%	27,27%
Instructor8	100,00%				Instructor51	100,00%			
Instructor9	50,00%			50,00%	Instructor52	100,00%			
Instructor10	85,71%			14,29%	Instructor53	88,24%	11,76%		
Instructor11	100,00%				Instructor54	92,31%	7,69%		
Instructor12	87,50%	12,50%			Instructor55	71,43%	14,29%		14,29%
Instructor13	80,00%			20,00%	Instructor56	86,67%	6,67%		6,67%
Instructor14	62,50%	12,50%		25,00%	Instructor57	100,00%			
Instructor15		100,00%			Instructor58		100,00%		
Instructor16	77,78%	11,11%	11,11%		Instructor59	18,18%	18,18%	27,27%	36,36%
Instructor17	100,00%				Instructor60	62,50%	6,25%		31,25%
Instructor18	83,33%	8,33%	8,33%		Instructor61	100,00%			
Instructor19	72,73%	4,55%	18,18%	4,55%	Instructor62	84,21%	10,53%	5,26%	
Instructor20	50,00%	8,33%	16,67%	25,00%	Instructor63	66,67%	16,67%	11,11%	5,56%
Instructor21	85,71%	7,14%		7,14%	Instructor64	66,67%	33,33%		
Instructor22	59,09%	22,73%	18,18%		Instructor65	20,00%	10,00%	30,00%	40,00%
Instructor23	78,26%	13,04%		8,70%	Instructor66	84,62%	7,69%		7,69%
Instructor24		100,00%			Instructor67	52,94%	17,65%	17,65%	11,76%
Instructor25	83,33%		16,67%		Instructor68	50,00%		50,00%	
Instructor26	71,43%		14,29%	14,29%	Instructor69	85,71%	14,29%		
Instructor27	82,61%	13,04%		4,35%	Instructor70	100,00%			
Instructor28	100,00%				Instructor71		100,00%		
Instructor29	60,00%	10,00%	10,00%	20,00%	Instructor72	100,00%			
Instructor30	83,33%		16,67%		Instructor73	66,67%			33,33%
Instructor31	91,30%			8,70%	Instructor74	85,71%	14,29%		
Instructor32	80,00%			20,00%	Instructor75	92,86%	7,14%		
Instructor33	43,75%	18,75%	6,25%	31,25%	Instructor76	100,00%			
Instructor34	100,00%				Instructor77	100,00%			
Instructor35	33,33%			66,67%	Instructor78	30,00%	35,00%	15,00%	20,00%
Instructor36	100,00%				Instructor79	64,29%	14,29%	14,29%	7,14%
Instructor37	100,00%				Instructor80	60,00%		40,00%	
Instructor38	100,00%				Instructor81	66,67%			33,33%
Instructor39	92,31%	7,69%			Instructor82	100,00%			
Instructor40	64,29%	14,29%	14,29%	7,14%	Instructor83	100,00%			
Instructor41	100,00%								
Instructor42	100,00%								
Instructor43	90,00%		10,00%						

The third aim of this study was to present a descriptive examination regarding the students across the groups. For this purpose, students were first assigned to their classes and

then frequencies of students were calculated per teacher (see Table 4). The purpose of this was to present the method rather than the evaluation of the teacher, thus only a few selected examples are presented here. To illustrate, 66% of the students who evaluated instructor 35 were placed in Cluster 4. As explained above, although the students in this group had positive ratings about their instructors, they might have felt unrewarded. Instructor 46 is also a good example of a descriptive examination of student evaluations. It can be interpreted that overall, 66% of the students who evaluated instructor 46 expressed negative opinions (see Table 4). However, some individual students appeared very satisfied with the instructor's teaching activities. This tendency is also seen in the ratings of instructors: 3, 8, 11, 17, 28, 36, 37, 38, 51, 52, 57, 61, 70, 76, 77, 82, and 83 (see Table 4).

As indicated in many studies, typical methodological strategies assume the population is homogeneous. However, this study aimed to identify distinct groups, therefore LCA was used. If the group in this study were assumed to be homogeneous, the analysis would simply be carried out based on the means of student evaluation of teacher. By applying this analysis, different aspects of student attitudes towards their teachers were examined and four main tendencies- the positive, undecided, negative, unrewarded- were found. These findings appear to be consistent with a number of other studies conducted on SET, all of which indicate the importance of instructor- level factors, such as use of time, concern for students' performance, interaction with students, and the ability to motivate students (Boex, 2000; Pozo-Munoz et al., 2000). In addition to these, some studies highlighted the importance of factors relating to personality, such as a caring nature and interpersonal skills (Brown & Atkins, 1993; Patrick & Smart, 1998). These findings are supported by the analysis of Cluster 4, which seems to show that positive perception of teaching ability can be offset by a perceived lack of praise, and indicates the importance of instructor personality issues, such as showing a caring attitude and valuing student contribution.

The results of this study also appear to be consistent with the results of other SET studies regarding student-level and subject-level factors, such as reasons for taking the course, gender, age, the amount of knowledge they gained (Basow, 2000; Marsh, 1987). Some SET studies indicate that females are more likely to give positive ratings than males and students who gain knowledge show their favoritism by giving higher ratings (Marsh, 1987; Mason et al., 1995). The results of this study showed that females gave higher ratings to their instructors' in-classroom teaching activities although some felt unrewarded and not praised. A substantial amount of research into SET shows that students give higher ratings if they gain knowledge, and if the purpose of the evaluation is explained to the students (Basow, 2000; Marsh, 1987). The majority of students in this study gave high ratings to their instructors; possibly reflecting the clear explanations given during the orientation week, and the amount of knowledge gained by the end of the academic year, when the SET was carried out. Some studies on SET also highlighted the importance of the subject itself, class-size, and method of administration. In this study, the class sizes were between 15-18, the subject was English language, and the questionnaires were filled out anonymously. The higher ratings given were consistent with the results of previous research which indicates higher ratings based on subject (Humanities and Arts are rated more positively), class size (smaller classes give

higher ratings) and anonymity (Braskamp & Ory, 1994; Davies et al., 2006; Fries & McNinch, 2003; Neumann, 2000; Worthington, 2002).

4. Conclusion and implications

Students' evaluation of teaching and teacher has been focus of many studies for years, however, the analysis of the data in these studies has mainly based on typical methodologies. Traditional Latent Variable models used in the data analysis, like factor analysis, are variable oriented, while LCA is a person- oriented method which allows the identification of distinct groups in heterogeneous samples, as in this study. To this end, this study has attempted to identify distinct groups of students based on their SET ratings and whether there is a discrepancy in the identified classes in terms of level of success and gender using the LCA method. The method also helped to present a descriptive examination regarding the students who evaluated the instructors and which classes they belonged to. The results of the study indicated that following three conclusions can be drawn from this research; that different typological structures of student exist in the institution, that there are differences in the identified classes in terms of gender, and that, regardless of whether they were successful or not, students were generally positive about teachers.

The analysis of the study suggests that students agree the teachers are able to present their abilities well, although some feel unrewarded, and thus, the LCA method in this study has led the institution to offer opportunities for teachers to raise awareness of the benefits of praising students' positive behaviours. Based on the results of the study, it could be suggested that LCA method can be more often used in evaluation research, specifically in identifying distinct groups in education. Following this, longitudinal studies could be conducted in order to identify any changes by using latent transition analysis, which is another type of LCA. It is believed that using these methods in evaluations will provide researchers with deeper insight.

In conclusion, although SET remains controversial, it is important to note that these evaluations certainly provide valuable information where appropriate methods are employed.

References

- Basow, S.A. (2000). Best and worst professors: Gender patterns in students' choices. *Sex Roles: A Journal of Research*, 43(5), 407-417. <http://dx.doi.org/10.1023/A:1026655528055>
- Boex, L.F.J. (2000). Attributes of effective economics instructors: an analysis of student evaluations. *The Journal of Economic Education* 31(3), 211-227.
- Braskamp, L. A., & Ory, J.C. (1994). *Assessing faculty work: enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Brown, G., & Madeleine, A. (1993). *Effective Teaching in Higher Education*. London: Routledge.
- Cashin, W.E. (1990). Students do rate different academic fields differently. In M. Theall & J. Franklin (Eds.), *Student Ratings of Instruction: Issues for Improving Practice*. New

- Directions for Teaching and Learning, (pp. 113-121). San Francisco: Jossey-Bass.
- Collins, L. M., & Lanza, S.T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health Sciences*. New Jersey: John Wiley& Sons, Inc., Publication.
- Davies, M., Hirschberg, J., Lye, J., & Johnston, C. (2010). A systematic analysis of quality of teaching surveys. *Assessment & Evaluation in Higher Education*, 35(1), 83-96. <http://dx.doi.org/10.1080/02602930802565362>
- Davies, W.M., Hirschberg, J., Lye J., Johnston C., & McDonald I. (2006). What Influences Teaching Evaluations? Evidence from a Major Australian University. *The Business Review, Cambridge*, 6(1), 146-152.
- Feldman, K.A. (1993). College students' views of male and female college teachers: Part II--evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34,151-211. <http://dx.doi.org/10.1007/BF00992161>
- Fries, C. J., & McNinch, R.J. (2003). Signed versus unsigned student evaluations of teaching: A comparison. *Teaching Sociology*, 31(3), 333-344.
- Germain, M.L., & Scandura, T.A. (2005). Grade Inflation and Student Individual Differences as Systematic Bias in Faculty Evaluations. *Journal of Instructional Psychology*, 32(1), 58-67.
- Jacobs, L. C. (2002). Student rating of college teaching: what research has to say. Retrieved 21.03. 2011 from http://www.iub.edu/~best/pdf_docs/student_ratings.pdf
- Jones, J. (1989). Students' ratings of teacher personality and teaching competence. *Higher Education*, 18, 551-558.
- Lowman, J., & Mathie, V.A. (1993). What should graduate teaching assistants know about teaching? *Teaching of Psychology*, 20, 84-88.
- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Mason, P.M., Steagall, J.W., & Fabritius, M.M. (1995). Student Evaluations of Faculty: A New Procedure for Using Aggregate Measures of Performance. *Economics of Education Review*, 14(4), 403-416.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. In J. A. Hagenaars, & A. L. McCutcheon (Eds.) *Applied latent class analysis* (pp.56-89). New York: Cambridge University Press.
- Millea, M., & P.W. Grimes. (2002). Grade expectations and student evaluation of teaching. *College Student Journal*, 36(4), 582-590.
- Murray, H.G., Rushton J.P., & Paunonen S.V. (1990). Instructor personality traits and student

- instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2), 250-261. <http://dx.doi.org/10.1037/0022-0663.82.2.250>
- Muthen, B., & Muthen L. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6), 882-891. <http://dx.doi.org/10.1111/j.1530-0277.2000.tb02070.x>
- Neumann, R. (2000). Communicating student evaluation of teaching results: rating interpretation guides. *Assesment & Evaluation in Higher Education*, 25(2), 121-134. <http://dx.doi.org/10.1080/02602930050031289>
- Nylund, K.L., Asparouhov T., & Muthen B.O. (2007). Deciding on the number of classes in latent class analysis in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.
- Ory, J.C., & K. Ryan. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P. Abrami, and L. Mets (Eds.), *The Student Ratings Debate: Are they Valid? How can we best Use Them?* New Directions for Institutional Research, 109, San Francisco: Jossey-Bass.
- Patrick, J., & R.M. Smart. (1998). An empirical evaluation of teacher effectiveness: the emergence of three critical factors. *Assessment & Evaluation in Higher Education*, 2(2), 165-178. <http://dx.doi.org/10.1080/0260293980230205>
- Pozo-Munoz, C., Rebolloso-Pacheco, E., & Fernandez-Ramirez, B. (2000). The 'ideal teacher' implications for student evaluation of teacher effectiveness. *Assesment & Evaluation in Higher Education*, 25(3), 253-263. <http://dx.doi.org/10.1080/02602930050135121>
- Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education. *Assesment & Evaluation in Higher Education*, 25(4), 397-405. <http://dx.doi.org/10.1080/713611436>
- Vermunt, J.K., & Magidson, J. (2005). *Latent GOLD 4.0 User's Guide*. Belmont, Massachusetts: Statistical Innovations Inc.
- Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*, 16, 12-14.
- Wolfer, T. A., & Johnson, M.M. (2003). Reevaluating student evaluation of teaching: The teaching evaluation form. *Journal of Social work Education*, 31, 111-121.
- Worthington, A. C. (2002). The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A Case Study in Finance Education. *Assessment & Evaluation in Higher Education*, 27(1), 49-64. <http://dx.doi.org/10.1080/02602930120105054>
- Zhang, L. (2004). Do university students thinking styles matter in their preferred teaching approaches? *Personality and Individual Differences*, 37, 1551-1564.

<http://dx.doi.org/10.1016/j.paid.2004.02.012>

Note

Note 1: Indicators were selected based on their R square values.

Copyright Disclaimer

Copyright reserved by the author(s).

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).