Dynamic clustering to evaluate satisfaction with teaching at university

Francesca Bassi, Department of Statistical Sciences, University of Padua, Italy, via C. Battisti 241, 35121 Padua Italy, francesca.bassi@unipd.it ph. ++300498274152 -

Abstract:
Purpose
In this paper, students' satisfaction with the didactics in a large Italian university, that of Padua, is measured, giving special attention to its evolution over time in consecutive academic years. The overall level of the quality of the didactics is examined and its change over time is modeled. Moreover, the effect of courses' and teachers' variables on it is estimated.
Methodology
Latent cluster lass models and mixture latent class Markov models are estimated in order to identify groups of courses that are homogeneous for the level of the quality of the didactics. Evolution over the three academic years of satisfaction is monitored. The effect on the clustering and its dynamics of potential covariates is also examined.
Findings
Results of model estimation reveal some interesting evidences that are important indications for the university management to define targeted strategies to elevate teaching quality.
Originality
The paper gives its original contribution both on the side of methods applied to analyze data collected with students evaluation of teaching and on the evidences obtained for a large university.

Keywords: quality of the didactics, university management, latent class models

Introduction

Customer satisfaction in the service market is strictly connected to quality, for this reason, a thorough comprehension of clients' opinion is extremely important in giving suggestions to improve any kind of offer. In the same way, in higher education, the service provider, the university, needs to know students' evaluation of teaching in order to measure and increase the quality of the didactics (Marzo Navarro et al., 2005; Nixon et al. 2016).
Attention to quality of teaching started at the beginning of 1990 and has had increasing importance in all universities since then; across European institutions it is ruled by the Bologna Process (Keeling, 2006). In this context, students' opinion is very important as it is shown by the direct involvement in the process of the European Students Union. Among the various instruments that can be used to evaluate the quality of the didactics, the so-called Student Evaluations of Teaching (SET) has a prominent role. This is a multi-item questionnaire that collects opinions on the various aspects of an university course (Zabaleta, 2007). Even if it is widely known that higher education institutions have to ask students' opinion on the didactics, there is still a debate both on how this information should be collected and especially used. Moreover, there is still recent literature discussing on the correct definition of "good teaching", on the need of complementary sources of information on quality of the didactics in addition to students' opinions and on the effect of this measure by external factors not directly linked to quality (Dalla Zuanna et al., 2015). Sporen (2010), for example, suggests that students' evaluation of teaching may be influenced by various elements, such as students' and teachers' personal characteristics, as well as by courses features. Many recent papers shown these evidences and measured magnitude and direction of influence, for students' characteristics, see, for example, (Beran and Violato, 2005), for teachers' variables, (Griffin, 2004), for courses' features (Bedard and Kuhn, 2008).
In this paper, students' satisfaction with the didactics in a large Italian university, that of Padua, is measured, giving special attention to its evolution over time in three consecutive academic years. The University of Padua was founded in 1222 and is one of the largest in Italy with around 61,000 students and 2,000 professors working in 32 different departments and 8 schools: Agricultural Sciences and Veterinary Medicine, Economics and Political Sciences, Law, Engineering, Medicine, Psychology, Science, Human and Social Science. The Italian university system is based on the so called 3+2 reform, which started in academic year 2001-2002 and it is organized in cycles with three consecutive

levels: a first-cycle academic degree lasting three years, a second-cycle academic degree of two years and a PhD course. There exists a small group of single-cycle degrees lasting five years, specifically in Medicine, Veterinary, Law and Architecture. (Maggiolaro et al. 2017). At the University of Padua there are 80 first-cycle (bachelor) degree courses, 84 second-cycle (master) courses and 9 single-cycle (5-year-long) degree courses. The University of Padua is ranked at top levels among Italian universities both for the didactic and research activity.

This university started to collect students' opinion in the academic year 1999-2000 and still continues, now by means of an online questionnaire. The survey about students' opinion supports the various levels of the internal evaluation process; the survey results are given in a detailed form to individual professors and managers of the various organizational structures (course councils, departments, athenaeum schools). Furthermore, some succinct results are published in aggregated form on the university website. Specifically, for each teacher and course, the following indicators are published: the overall level of satisfaction; an indicator related to the organizational aspects of the course (clarity of scopes, examination arrangements, observance of timetable and didactic material); and an indicator related to efficacy of didactics (interest stimulation and clear explanation). We dispose of a dataset recording students' evaluations over three consecutive academic years of courses that did not change teacher nor dimension together with some information on course and teacher characteristics. Estimating appropriate specifications of latent class models, groups of courses that are homogeneous for the level of the quality of the didactics are identified and evolution over the three academic years of satisfaction is monitored. The effect on the clustering and its dynamics of potential covariates is also examined. From the point of view of the university management, it is not only important to assess students' satisfaction but also to follow its evolution over time and to identify which factors, both on teachers' and courses' side may have an impact on it as a vast recent literature clearly shows (Wodall et al. 2012).

Results of model estimation reveal some interesting facts. The level of satisfaction is, on average not the same across the different schools of the university, nor in bachelor and master degrees. The number of students attending the course and the number of teaching hours may have a significant effect on perceived quality, the role of the teacher is, instead, not important. Over the three academic year some didactic activities improved their quality as from the students' judgment, however, there exist still a small group of courses that did not improve and even worsen their performance. All these evidences are important indications for the university management to define targeted strategies to elevate teaching quality.

The paper is organized as follows: in section 1 it is described how quality of the didactics is measured at the University of Padua with a special attention given to the questionnaire, its properties and the data at our disposal. In Section 2, latent class cluster models are introduced and the mixture latent class Markov model is specified. Section 3 lists results on models' estimation and Section 4 concludes.

1. Measuring quality of the didactics at the University of Padua

The main aim of all actions devoted to the measurement of students' satisfaction at the University of Padua is to evaluate the level of perceived quality on all didactic activities organized in one academic year in order to give to the teachers information useful to eventually improve their teaching and related services. Two main instruments are used: a questionnaire with a limited number of open questions on good and bad aspects of the course and eventual suggestions to the teacher for improvements, and a structured questionnaire that collects students' judgments on all aspects regarding an academic course on a scale ranging from 1 to 10, where 10 describes the optimal situation. This instrument has been slightly changed from its first adoption in the academic year 1999-2000 in order to improve the quality of the collected information and to follow suggestions given by the Italian Agency for University Evaluation (ANVUR). Since the academic year 2010-2011, it is proposed to the students through the web as a CAWI (Computer Assisted Web Interviewing) survey at the end of each course, when the students enroll for the exams.

In the academic year 2012-2013, the questionnaire presented to the students began with two introductory questions: the first one asked if the student was available to participate in the survey (if the student was not, no other question was posed), the second one asked what percentage of the

lessons of the course under judgement was attended by the student. If the student attended less than 30% of the lessons, he was asked to answer only to seven selected items and to a question on why he attended so few classes; otherwise, all 18 items were proposed. In the following, the 18 items composing the scale to measure student satisfaction in the case of more than 30% of classes attended are reported.

Item 01 At the beginning of the course, were aims and topics clearly outlined?
Item 02 Were examination arrangements clearly stated?
Item 03 Was classes timetable observed?
Item 04 Is the number of lessons adequate to the course program?
Item 05 Is preliminary knowledge sufficient to understand all topics?
Item 06 Does the teacher stimulate interest towards the topic?
Item 07 Does the teacher clearly explain?
Item 08 Is the suggested material for study adequate?
Item 09 Is the teacher available to the needs of the students?
Item 10 Was the teacher available during office hours?
Item 11 Are laboratories/practical activities/workshops, if included, adequate?
Item 12 Are classrooms adequate?
Item 13 Are rooms for laboratories/practical activities/workshops adequate?
Item 14 How much are you satisfied about this course?
Item 15 Is the requested workload proportionate to the number of credits assigned to the course?
Item 16 Independently on how the course was taught, how much are you interested in the topic?
Item 17 How much is the course consistent with the whole degree?
Item 18 Does the course prepare to work?

Students attending a master degree are requested to answer to the following additional items:

Considering your bachelor degree, say how much of the contents of this course are
Item M01 a repetition of what you already studied
Item M02 a deepening of contents already proposed
Item M03 totally new topics.

The University of Padua publishes on its webpage part of the information collected with the above questionnaire. Specifically, for each teacher and course, the following indicators are circulated: the overall level of satisfaction based on item 14; an indicator related to the organizational aspects of the course, obtained as the arithmetic mean of items 01 (clarity of scopes), 02 (examination arrangements), 03 (timetable) and 08 (observance of timetable); an indicator related to efficacy of didactics, obtained as the arithmetic mean of items 06 (interest stimulation), 07 (clear explanation), and 09 (availability to needs of the students). Starting from the subsequent academic year 2013-2014, item 09 was eliminated by the indicator. Bassi at al. (2017) showed that the scale and the two indicators of efficacy of didactics and of organizational aspects are valid and reliable.

For our analysis we could dispose of a much smaller sample: 1,847 didactic activities that were evaluated for three consecutive academic years (2012-2013, 2013-2014 and 2014-2015) and did not change teacher nor number of teaching hours in the reference period. Observations with missing data or evident errors were excluded from the analysis. Only questionnaires filled in by regular students who attended at least 50% of the lessons were taken into account. For each course we have also information on the type of degree: bachelor, master or 5-year-long, number of teaching hours and corresponding credits (ECTS), university school were the course is given, role of the teacher, whether assistant, associate, full professor or other. For reasons of privacy of the data, all information was anonymized: courses, schools and teachers were given a code, not their name. In our sample, 1,057 didactic activities are given in Bachelor degrees, 471 in Master degrees and the remaining in 5-year-long degrees. The average number of filled in questionnaires per didactic activity was 11.41, 10.62 and 11.19 in the three consecutive academic years.

Table 1 lists the means and the standard deviations for the 12 items proposed in all three academic years, the mean level of satisfaction over the 11 items (item 14 is left out), and the two indicators of satisfaction with organizational aspects (OA), that is obtained as an average of items 01, 02, 03 and 08,

and efficacy of didactics (ED), obtained combining items 06 and 07, over our sample of didactic activities. Students give lower scores to items 04 (preliminary knowledge) and 15 (workload), higher scores to items 03 (course timetable) and 10 (teacher availability). However, the distribution of judgments is very asymmetric with lower scores used only rarely. Comparing means over time shows a quite stable situation, only a slight decrease in satisfaction for almost all items may be detected. However, these are average values, over almost 2,000 evaluated activities, with non-negligible standard deviations. We are interested in finding out if there are courses with patterns of evolution over time of the judgments expressed by students very different from the average either because they show an increase or a decrease. The scope of this paper is to identify these particular didactic activities and possibly to understand which factors cause the specific dynamics over time.

Table 1 about here

We performed an exploratory factor analysis on data collected in the most recent academic year 2014-2015 to explore the latent structure of our data. We excluded from this analysis item 14 which measures overall satisfaction with the didactic activity and can be seen as a sort of summary measure of the other 11 items. Factor analysis is a commonly used statistical tool for describing the associations among a set of manifest variables in terms of a smaller number of underlying continuous latent factors (Bartholomew and Knott, 2011). One factor explains almost 81% of total variance and shows very high loadings, greater than 0.84, with all 11 items. This latent factor represents students' satisfaction with university courses. This result does not contradict the evidence reported in Bassi et al., 2017 about four latent dimensions for the measurement scale; in this analysis we could consider only the 11 selected items, out of 17, that were proposed to the students in three consecutive academic years.
Factor loadings and item-to-rest correlation coefficients are all greater than 0.8, Cronbach's Alpha coefficients when each item is deleted are listed do not exceed the value of the same coefficients for the complete scale (0.976)The figures confirm that the items that we are considering to measure student's satisfaction constitute a measurement scale with the property of reliability. Scale validity is ensured by the fact that the 11 items are all highly correlated (> 0.77) with item 14, measuring overall satisfaction, that is considered as the gold standard (see, for a more extended analysis of these items, Guerra et al., 2017).

A confirmatory factor analysis shows that there is a problem of fit. This is probably due to the fact that the model does not include co-varying error terms and 11 highly correlated items are too many for only one underlying factor. The 11 items were reduced to seven indicators in the following way. Items 01, 02, 03 and 08 were aggregated in the indicator that measures satisfaction with reference to organizational aspects (OA); items 06 and 07 were aggregated in the indicator that measures satisfaction with reference to efficacy of didactics (ED). As already said, these two indicators are published by the University of Padua for every didactic activity and have shown to be valid and reliable in a previous work (Bassi et al., 2017). Items 04 (preliminary knowledge), 10 (availability during office hours), 11 (laboratories), 15 (workload) and 16 (interest in the topic) complete the set of indicators. Performing factor analysis with these seven indicators shows the presence of one underlying factor that explains more than 80% of total variance and factor loadings are all greater than 0.82. Moreover, the fit of the model with the reduced number of indicators improves with reference to several indexes (AIC, BIC, likelihood ratio statistics, Root Mean Squared Error). A Cronbach's alpha coefficient of 0.946 ensures internal reliability. The rest of the paper is based on this summarized set of items into seven indicators.

2. Latent class cluster models

Latent class models belong to finite-mixture modeling and were introduced by Lazarsfeld (1950) as a method to identify heterogeneity in a group of respondents to a survey. Latent class models constitute a model-based approach to clustering (Magidson and Vermunt, 2002); its advantages over cluster

analysis are due to the fact that belonging of each unit to a group is given by a probability estimated via maximum likelihood and not by a simple indicator of distance among units and groups. Moreover, latent class modeling is at the same time more flexible in treating variables measured on different scales and more rigorous in providing measures of fit to identify the optimal number of groups. Cluster analysis suffers instead of lack of robustness: different measures of distance, algorithms, criteria to determine the optimal number of groups may lead to different solutions.
A latent class cluster (LCC) model (Vermunt and Magidson, 2002) has the form in equation (1)

$$f(\mathbf{y}_i \mid \mathbf{z}_i, \boldsymbol{\theta}) = \sum_{k=1}^{K} P(x = k \mid \mathbf{z}_i) \prod_{j}^{J} f_k(y_{ij} \mid x, \mathbf{z}_{ij}, \boldsymbol{\theta}_k) \quad (1)$$

where $\boldsymbol{y}_i$ denotes the values observed for unit i on a set of J continuous indicators, x is a unobservable categorical variable with K classes, $\boldsymbol{\theta}_k$, the unknown parameters of the specific density k, $\boldsymbol{z}_i$ the vector containing the values of R covariates observed on unit i. $P(x=k|\boldsymbol{z}_i)$ is the probability of belonging to latent class k given individual covariates. $f(y_{ij}|x,\boldsymbol{z}_{ij},\boldsymbol{\theta}_k)$ is the probability density of each observed variable that depends on the cluster an on individual covariates. The distribution of $\boldsymbol{y}_i$, given the unknown parameters is assumed to be a mixture of class-specific densities; this distribution is assumed to depend also on observed covariates.
In this application, since all observed variables are continuous, they are assumed to be distributed as multivariate Normal conditionally on cluster belonging. As it is usual in latent class analysis, local independence is assumed.
A latent class Markov (LCM) model describes transitions among latent states (clusters) over time by means of a Markov chain (Bartolucci et al., 2013). A LCM model with a first order chain and covariates is reported in equation (2)

$$f(\mathbf{y}_i \mid \mathbf{z}_i, \theta) = \sum_{k_1=1}^{K} \dots \sum_{k_T=1}^{K} P(x_1 = k_1 \mid \mathbf{z}_{i1}) \prod_{t=2}^{T} P(x_t = k_t \mid x_{t-1} = k_{t-1}, \mathbf{z}_{it}) \prod_{j=1}^{J} \prod_{t=1}^{T} f_k(y_{ij} \mid x, \mathbf{z}_{ij} \boldsymbol{\theta}_k) \quad (2)$$

where $P(x_1=k_1|z_{i1})$ is the initial state probability which may depend on time-varying and time-constant covariates; $P(x_t=k_t|x_{t-1}=k_{t-1},z_{it})$ are transition probabilities which may depend on time-varying covariates. This model has been used in recent literature to study dynamic segmentation, see, for example Bassi (2017).
Finally, a mixed latent class Markov (MLCM) model (van de Pol and Langeheine, 1990) includes a mixture variable s with H categories that accounts for potential heterogeneity in the dynamics over clusters as in equation (3)

$$f(\mathbf{y}_i \mid \mathbf{z}_i, \theta) = \sum_{h=1}^{H} P(s = h \mid \mathbf{z}_i) \sum_{k_1=1}^{K} \dots \sum_{k_T=1}^{K} P(x_1 = k_1 \mid s, \mathbf{z}_{i1}) \prod_{t=2}^{T} P(x_t = k_t \mid x_{t-1} = k_{t-1}, s, \mathbf{z}_{it}) \prod_{j=1}^{J} \prod_{t=1}^{T} f_k(y_{ij} \mid x, s, \mathbf{z}_{ij} \boldsymbol{\theta}_k)$$

(3)

where $P(s=h|\boldsymbol{z}_i)$ are class proportions which may depend on time-constant covariates. All other elements that appear in equation (2), in the MLCM model may depend on s. A special specification of the MLCM model is the mover-stayer were H=2 and in one of the two latent chains the transition matrix is imposed to be equal to the identity matrix, which means that no changes across clusters are possible over time.

3. Results

In order to identify how many clusters can appropriately classify the courses at the University of Padua according to the judgments expressed by students, a LCC model was estimated on the data collected in academic year 2012-2013, obtaining the best fit with four categories of the latent variable

according to various criteria: AIC3, BIC, CAIC and percentage of classification error (van Kollenburg et al, 2015). Models were estimated with Latent GOLD (Madigson and Vermunt, 2005). Observed variables of the model are the seven indicators introduced in Section 2.
Table 2 reports clusters' sizes and the average value in each cluster per item. It appears quite evident that the four classes represent groups of courses taught at the University of Padua that received significantly different judgments by the attending students.. The level of satisfaction increases from cluster 1 to cluster 4. In cluster 1, we find courses that received an average judgment just above the minimum level of acceptable satisfaction, this seems to be the group of didactic activities where it is necessary to operate in order to increase quality. A second group of courses (34% of total) shows and average level of satisfaction above 7 on our scale; the third group (35%) shows an average satisfaction level greater than 8 and, finally, the cluster analysis identifies a group of courses judged to be excellent by the students (14%).
The covariates introduced in the LCC model that resulted significant in identifying clusters are the type of degree where the course is inserted, the school, the number of teaching hours and the number of filled-in questionnaires, that is a proxy of the number of students attending the course. Table 3 describes the four clusters with reference to the significant covariates in terms of conditional probabilities for categorical variables and of mean values for continuous variables. Master courses tend to be associated with cluster 4 (highest satisfaction), bachelor courses with clusters 1, 2 ad 3; students of longer courses, more teaching hours, tend to be more satisfied; the effect of the number of questionnaires is that the smaller classes tend to be associated with the lowest and the highest level of satisfaction; finally, there is an effect also of the university school which needs further investigation, clear evidences are that didactic activities taught in school 4 have a higher probability to be in cluster 1 where satisfaction is at the lowest level, while didactic activities from school 8 are more likely to be in cluster 4, where satisfaction is the highest. The same model, estimated on the data collected in the subsequent academic years, gives very similar results. An interesting evidence regards the fact that the dimension of cluster 1 decreases over the three academic years to 15.5% indicating that the number of critical didactic activities diminishes but, at the same time in this group the average level satisfaction decreases, indicating that there is a small group of courses in which quality at least does not improve.
This preliminary analyses suggest to the university management that there is one school in which students perceive the lowest level of the quality of the didactics. Also the evidence that students in the firs-level degree are less satisfied should stimulate a research of the causes of this phenomenon, that might be on courses and teachers side but also on other facts related to the students and their short experience with the higher education environment.
The main focus of this paper is however on the longitudinal dimension of satisfaction as the following analyses show.

Table 2 about here

Table 3 about here

At this point, it becomes interesting to study how satisfaction evolved over time in the three consecutive academic years, especially if some improvement, or some worsening, has been noticed by the students. In order to evaluate this problem, a LCM model was estimated with a first-order latent chain and observed variables being the seven indicators summarizing students' satisfaction in the three consecutive academic years for the sample of 1,847 courses. The best fitting model is LCM model with four latent states and a stationary chain. Transitions probabilities are imposed to be the same over time in order to obtain an identifiable model and for reasons of parsimony; moreover, a conditional test with the LCM model with heterogeneous transitions led to accept the hypothesis of a stationary chain. Latent states represent courses with similar level of students' satisfaction, their dimensions and profiles are consistent with the results obtained estimating the LCC model on the data collected in the first academic year of our reference period: there is a small group of courses where quality is judged to be very low (cluster 1), another small group where quality is, at the opposite, very high (cluster 2), and two other groups with medium (cluster 2) and high (cluster 3) satisfaction. Estimated transition probabilities (Table 4) show that there is a non-negligible dynamics over time across clusters, 54% of courses were judged differently during the observational period, from one academic year to the next

one, by the students. Another interesting evidence is that not all transitions show an improvement in the quality of the didactics, this is represented by the probabilities reported in the cells in the portion of the table under the diagonal; for example, 10% of course rated in cluster 4 at the beginning of the period, moved to cluster 1 where satisfaction is very low in the subsequent academic year.

Table 4 about here

The effect of the available covariates was estimated both on the probability of belonging to the latent state and on transition probabilities. The number of filled-in questionnaires, being a time-varying variable, was inserted only as a possible covariate for latent transitions. The degree in which the course is offered and the school resulted significant in affecting the initial state, the number of questionnaires, the role of the teacher and the school resulted significant in estimating transitions. Specifically, courses in bachelor degrees show a significant positive association with cluster 1, courses in master degrees have a significant positive association with cluster 4, and courses in 5-year-long degrees with cluster 2. The pattern of the significant effects of school is more complex: summarizing, schools 4, 5 and 7 are positively associated to cluster 1 where satisfaction is the lowest, school 8 is positively associated with cluster 4 where satisfaction is the highest. Unfortunately, for reasons of privacy, we cannot identify the schools, however the university management can benefit from this results since it knows how the school codes were assigned. For what concerns transitions over time, they are significantly influenced by the number of teaching hours and the number of filled-in questionnaires, however, this effect is asymmetrical.
A mover-stayer latent class Markov (MLCM) isolates the group of courses that did not change cluster and helps in looking more deeply in what happens to the didactic activities that were judged differently over time by the students. The best fitting model, in this case, is a mover-stayer model with four latent states, a first-order Markov chain for the movers and covariates. Its fit improves over the simple LCM model.
The latent states of the first-order Markov chain contain courses with similar level of students' satisfaction, the fours clusters are consistent in both classes with those obtained in the previous analysis; a smaller group of courses (12%, the so-called stayers) do not change cluster over the three academic year; this means that students judge them in a similar way, the remaining 88% of didactic activities change cluster over time. As it is clear from Table 5, satisfaction in the two classes reaches good average levels of satisfaction, with courses in the stayers category showing slightly higher levels; the only covariate that has a significant effect on classes is the school, specifically, in class 1 (stayers) we find a higher proportion of courses given in school 3.

Table 5 about here

Table 6 about here

For both classes, clusters represent groups of courses with a perceived different level of quality by students: in cluster 1 satisfaction is the lowest, the highest in cluster 4. The distribution of didactic activities in the four clusters is quite similar in the two classes, however the general level of satisfaction is higher in all courses belonging to the class of stayers: courses that do not change cluster over the three academic years are judged to have higher quality by students. The model identifies a small group of courses (222) with an average good level of satisfaction that do not change this level over time. The only critical subgroup is that of those courses (34) that are classified in cluster 1 and do not exit from it; they are concentrated in bachelor degrees of schools 5 and 7.
The estimated conditional probabilities in the transition matrix for the courses following the mover chain (Table 6) show that only 35% of the didactic activities remain in cluster 1, where satisfaction is the lowest in the subsequent academic year, the others move to a cluster where satisfaction is higher. On the other hand, 18% of courses that belong in the first academic year to cluster 2, where satisfaction is the medium, moved to cluster 1, as 36% of courses form cluster 3 and 62% of course in cluster 4 move to clusters were satisfaction is lower.

The effect of the covariates is very similar to that seen in the simple LCM model. The type of the degree in which the course is proposed and the school have a significant effect on the probability of the initial state: courses in bachelor degrees show a significant positive association with cluster 1, course in master degrees have a significant positive association with cluster 4, and courses in 5-year-long degrees with cluster 2. Courses in schools 4, 5 and 7 are positively associated with cluster 1, courses in school 8 with cluster 4. The type of degree, the number of teaching hours and the number of filled-in questionnaires have a significant impact on transitions: courses in bachelor degrees have higher probabilities to move to lower levels of satisfaction, the contrary is true for courses in master degrees and 5-year long degrees. The longer the course the higher the association with transition towards less satisfaction over time, same for the number of filled-in questionnaires which is a proxy of the number of attending students.

4. Concluding remarks

Scope of this paper is to analyze students' satisfaction at the large university, that of Padua considering also its evolution over a period of time.
The available data refer to a sample of almost 1,900 courses that were taught by the same teacher, giving the same number of credits, in three consecutive academic years, evaluated by the students who attended at least 50% of classes. Students were asked to fill in a web questionnaire with items on a 10-point scale. Only items proposed to the students in all three years were considered and summarized in a set of seven indicators for reasons of parsimony. These seven indicators constitute a valid and reliable scale of students' satisfaction.
Estimation of a latent class cluster model classifies the courses in four groups with an increasing level of perceived quality of the didactics. A mixture latent class Markov model studied the evolution over time of these courses. Two unobservable classes of didactic activities are identified: a small group of courses with an average higher level of satisfaction reported by attending students; these courses do not show a significant dynamic pattern over time, in the sense that they are stayers, the estimated transition matrix is the identity matrix. From a substantive point of view, this means that in these courses no change in the quality of the didactics was perceived by the students in subsequent academic years, however, in this group, the average level of satisfaction is high. In the larger group of courses, the transition matrix is estimated significantly different from the identity one, indicating that a percentage of courses has received a general global judgment that may change over time, although not always towards an improvement.
There is a clear need for the university management, not only to assess students' satisfaction but also to increase it. As for other services, satisfaction is a driver of loyalty and of good reputation 0t continue enroll new students in the subsequent academic year. Moreover, quality of the didactics is an indicators in national and international rankings of institutions that provide higher education. Specifically, in Italy, the assessment of students' satisfaction is mandatory by the National Agency of University Evaluation (ANVUR).
The results of model estimation can give important information to the university to improve quality, especially because they underline some critical points:
1) There is a non-negligible group of didactic activities (around 19% of total) that the students judge to lower their teaching quality in subsequent academic years.
2) Lower quality courses are more concentrated in bachelor degrees and in some schools, such as those numbered with 4, 5 and 7; we are not allowed to know the school names but this information is available to the university management.
3) Higher quality courses are more concentrated in school 8.
4) Courses in the bachelor degrees and in 5-years-long degrees tend to be more associated with the probability of moving to lower levels of students' satisfaction over time.
5) Longer courses, with more teaching hours, have a higher risk to move to lower satisfaction levels over time.
6) Courses with more attending students have a higher risk to move to lower satisfaction levels over time.

7) The role of teacher does not have a significant impact on the level of satisfaction nor on its dynamics over time.

Starting from the above evidences and with the possibility of knowing how school names were anonymized, the university management can implement targeted policies to work on critical courses. The is a rich literature that shows the link between students' satisfaction and students' retention at higher education institutions (see, for example, Browne et al., 1998 and Elliott, 2002). Enhancing students' satisfaction is an important goal of modern universities management (Marzo Navarro et al., 2005). In this context following satisfaction over time, identifying bad performing courses and potential instruments to afford critical situations is of strategic importance.

As usual, further analyses might be very useful, especially on the teachers' side. In this exercise, we could consider only the teacher role which did not result significant in determining the level of students satisfaction nor on its dynamics, however there might be other teacher characteristics, such as age, experience, use of specific didactic tools that may be instead very important for determining teaching quality.

Acknowledgment: Research for this paper was supported by grant BIRD162088/16 financed by the University of Padova for the project with title "Advances in Multilevel and Longitudinal Modelling".

References

Bartholomew, D. and Knott, M. (2011), *Latent Variable models and Factor Analysis, 3rd ed.*, Wiley, New York.

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2013), *Latent Markov Models for Longitudinal Data*, CRC Press Boca Rato, FL.

Bassi, F. (2017), "Longitudinal models for dynamic segmentation in financial markets", *International Journal of Bank Marketing*, Vol. 35, pp. 431-446.

Bassi F., Clerici, R., and Aquario, D. (2017). Students' evaluation of teaching at a large Italian university: validation of measurement scale, *Electronic Journal of Applied Statistical Analysis*, 10, 93-117.

Bedard, K. and Kuhn, P. (2008), "Where class size really matters: class size and students' ratings of instructor effectiveness", *Economics of Education Review*, Vol. 27, pp. 253-265.

Beran, T. and Violato, C. (2005), "Ratings of university teacher instruction: How much do student and course characteristics really matter?", *Assessment and Evaluation in Higher Education*, Vol. 30, pp. 593–601.

Browne, B.A., Kaldenberg, D.O., Brown, W.B. and Brown, D. (1998), "Student as customer: Factors affecting satisfaction and assessments of institutional quality", *Journal of Marketing for Higher Education,* Vol. 8, pp. 1-14.

Dalla Zuanna, G., Bassi F., Clerici, R., Paccagnella, O., Paggiaro, A., Aquario D., Mazzuco C., Martinoia, S., Stocco, C. and Pierobon, S. (2015), "Tools for teaching assessment at Padua University: role, development and validation", Report of Research Unit n.3 PRODID Project (Teacher professional development and academic educational innovation), Department of Statistical Sciences, University of Padua, Padua.

Elliott, K.M. (2002), Key determinants of students' satisfaction", *Journal of College Students Retention: Research Theory and Practice,* Vol. 4, pp. 271-279.

Marzo Navarro, M., Pedraja Iglesias, M. and Rivera Torres, P. (2005), "A new management element for universities: satisfaction with the offered courses", *International Journal of Educational Management,* Vol. 19, pp. 128-139.

Meggiolaro, S., Giraldo, A. and Clerici, R. (2017), "A multilevel competing risks model for analysis of university students' carrers in Italy", *Studies in Higher Education*, Vol. 42, pp. 1259-1274.

Griffin, B.W. (2004), "Grading leniency, grade discrepancy, and student ratings on instruction", *Contemporary Educational Psychology*, Vol. 29, pp. 410-425.

Guerra , M., Bassi, F. and Dias, J.G. (2017), "A decision support system to track courses with low-quality teaching", submitted.

Keeling, R. (2006), "The Bologna Process and the Lisbon Research Agenda: the European Commission's expanding role in higher education discourse", *European Journal of Education*, Vol. 41, pp. 203-223.

Lazarsfeld, P. (1950), "The logical and mathematical foundation of latent structure analysis", *Measurement and Prediction*, Vol. 4, pp. 362-412.

Magidson, J. and Vermunt, J.K. (2002), "Latent class models for clustering: a comparison with k-means", *Canadian Journal of Marketing Research*, Vol. 20, pp. 36-43.

Magidson, J. and Vermunt, J.K. (2005), "A nontechnical introduction to latent class", *Canadian Journal of Marketing Research*, Vol. 20, pp. 36-43.

Marzo Navarro, M., Pedraja Iglesias, M. and Rivera Torres, P. (2005), "A new management element for universities: satisfaction with the offered courses", *International Journal of Education Management*, Vol. 19, pp. 505-526.

Nixon, E., Scullion, R. and Hearn, R. (2016), "Her majesty the student: marketised higher education and the narcissistic (dis)satisfaction of the student-consumer", *Studies in Higher Education*, DOI: 10.1080/03075079.2016.1196353.

Spooren, P. (2010), "On the credibility of the judge. A cross-classified multilevel analysis on student evaluations of teaching", *Studies in Educational Evaluation*, Vol. 36, pp. 121–131.

Van de Pol, F. and Langeheine, R. (1990), "Mixed Markov latent class models", *Sociological Methodology*, Vol. 33, pp. 23-247.

Van Kollenburg, G.H., Mulder, J. and Vermunt, J.K. (2015), "Assessing model fit in latent class analysis when asymptotics do not hold", *Methodology*, Vol. 11, pp. 65-79.
Vermunt, J.K., Magidson, J. (2002), "Latent class cluster analysis", in J.P. Hagenaars and A. McCoutcheon (Eds.), *Applied Latent Class Analysis*, Cambridge university Press, Cambridge, pp. 89-106.
Woodall, T., Hiller, A. and Sheilagh, R. (2012), "Making sense of higher education: students as consumers and the value of the university experience", *Studie in Higher Education*, Vol. 39, pp. 48-67.
Zabaleta, F. (2007), "The use and misuse of student evaluations of teaching", *Teaching in Higher Education*, Vol. 12, pp. 55-76.