

# Looking for patterns of change amid pandemic period in students' evaluation of academic teaching

Annalina Sarra<sup>1†</sup>, Adelia Evangelista<sup>1\*†</sup>, Barbara Iannona<sup>1†</sup>  
and Tonio Di Battista<sup>1†</sup>

<sup>1\*</sup>Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University "G. d'Annunzio", Viale Pindaro,42, Pescara, 65127, Italy.

\*Corresponding author(s). E-mail(s): [adelia.evangelista@unich.it](mailto:adelia.evangelista@unich.it);

Contributing authors: [annalina.sarra@unich.it](mailto:annalina.sarra@unich.it);  
[barbara.iannone@unich.it](mailto:barbara.iannone@unich.it); [tonio.dibattista@unich.it](mailto:tonio.dibattista@unich.it);

<sup>†</sup>These authors contributed equally to this work.

## Abstract

The COVID-19 pandemic has shaken not only the global economy but every development field, including all levels of the education sector and in every place in the world. The wide spread of this pandemic disaster has undoubtedly changed the education landscape worldwide. Online teaching and learning become the primary instruction method and the global world of schools, colleges and universities were forced to adapt this model. The first concern about online learning is whether this method is effective compared to traditional face-to-face lessons. In this paper, we carried out a quantitative analysis to explore variations in university students' feedback on learning experience in the context of this new challenging situation caused by the current COVID-19 pandemic. By adopting an IRT modeling, we compared the appreciation of some aspects of 41 courses taught at the University of Chieti-Pescara (Italy) during the educational emergency with that of the previous year. Overall, from the results of this study, it is arisen that students have given positive feedbacks on their learning experiences and their effectiveness.

**Keywords:** IRT, Linear Logistic Models with Relaxed Assumptions, Higher Education, Online Learning, Learning Satisfaction Data

# 1 Introduction

The coronavirus disease (COVID-19) spread rapidly worldwide and it was declared, by the World Health Organization, as a pandemic phenomenon on March 11, 2020. The COVID-19 pandemic brought a global crisis, and its consequences to date have been severe and significant on all sectors of our lives, especially on health and economy, causing, among others, acute economic losses and a critical pressure on health systems. Response to the pandemic has been widely different from country to country, depending on the alignment to the national government and their pre-existing investment in resilience. In the educational context, the immediate impact of COVID-19 has been lockdown and the enforced closure of educational institutions. Schools, colleges and universities have been shut down across the world and forced to implement a transition from traditional face-to-face teaching methods to online teaching or a combination of online and traditional teaching (blending), to ensure continuity of education [2, 31, 32, 82]. During the COVID-19 pandemic, e-learning, as a distance learning strategy, enabled universities to use a variety of online learning platforms; Zoom, Google classroom, Microsoft Teams, Blackboard are some of the most adopted versatile platforms for online learning and teaching processes during the ongoing outbreak. According to recent reports, COVID-19 has caused an education crisis (OECD, 2020) [41]. Different published works have already documented that the emergency of online learning has resulted in many difficulties for teachers, students and higher education administrators [11, 43, 83]. While some nations, thanks to their economic and human resources, organizational skills and students' ability to interact with the electronic environment, have been able to easily shift to online learning, for others the transition from face-to-face to online teaching was fraught with difficulties. The limited availability to computer and internet resources, and thus some problems in accessing online lessons, materials downloading, the lack of digital skills in using online platforms, are the main challenges and obstacles encountered by the students. Other studies have tried to examine the feedbacks of universities to this new challenge all over the world [38, 39, 54].

For instance, a research carried out by [10] outlined the reaction of several universities through countries and demonstrated diverse responses to the complex challenge. Overall, the ability of colleges and universities to quickly shift toward online and virtual courses is impressive. At the same time, the effects on teaching and learning has been mixed [17, 46, 75, 79–81].

Additionally, further comparative surveys were aimed at exploring whether face-to-face or traditional teaching methods are more productive or whether online or hybrid learning is better [34, 47, 65]. Findings arising from these studies give evidence of a better students' performance in online learning than in traditional one. Besides, in the past, several researches have been carried out to explore students' satisfaction with online learning teaching, which was used as a pedagogical method for many years in various fields [7, 53, 61, 63, 74, 78].

In this regard, it is worth noting that the both proponents and opponents of online learning continued, over time, the discourse surrounding quality of online learning provisions, student engagement and satisfaction [9, 33, 35, 37].

However, a limited literature is available on students' satisfaction with online experiences during the coronavirus pandemic [66]. To fill this gap, in this paper, we explore variations in students' feedback on learning experience in the context of this new challenging situation of digital learning caused by COVID-19 disease. Using quantitative data obtained through the administration of the same questionnaire to face-to-face and online learners, this research compares the appreciation of courses of University of Chieti-Pescara (Italy) during the educational emergency with that of previous year. We address this research question through the Item Response Theory (IRT) modelling techniques. IRT [21], successfully applied in many settings, refers to a family of probabilistic models that attempt to explain the relationship between latent traits (e.g. attitudes, abilities, skills, satisfaction) and their manifestations. IRT assumes that the latent construct and the items of a measure are organized in an unobservable continuum. In our context, we exploited the attractive features of a particular class of IRT models, known as Linear Logistic Models with Relaxed Assumptions [28], to detect patterns of change amid pandemic period in students' evaluation of academic teaching compared to the previous year. The remainder of this paper is organized as follows. Section 2 is devoted to illustrate the challenges that universities are facing with over the last years, and the importance of the assessment of quality within higher institutions. Section 3 covers the basics of IRT models adopted in this study to capture the change amid pandemic period in students' evaluation of academic teaching. Data and main results are then presented in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper.

## **2 Background: quality within higher education institutions**

Universities, both nationally and internationally, have experienced a profound evolution over time, often under the pressure of reforms desired by individual States [1, 6, 24, 44, 56]. In particular, in the last years, many universities are in front of important challenges, regarding the decline of enrolments, an increased competition, the low number of graduates in the labour force, high rate of dropout other than drastic financial troubles. Furthermore, in the new context, the funding State calls for quality services by setting specific targets. Various indicators are in fact mirrored in funding formulae and public funding is delivered to universities according to different parameters [64]. The ways higher institutions are organized, financed and delivered are at the basis of their complexity. Universities are open systems that establish relationships with the environment. Such relationships, in terms of constraints and opportunities, must be considered as a strategy to creating public value [55, 69]. Over

the past decades, quality of teaching and learning has become a major strategic issue in tertiary education systems across the globe [22]. Universities are increasingly aware of the importance of students' satisfaction for reputation as well as for educational reasons [60]. The perceived quality of the educational experience is primarily important as diagnostic feedback to administrators and faculty for improving teaching; in addition, it provides information for students for the selection of courses and instructors. Mapping the quality of teaching is necessary, at least in principle, also for the government bodies to modulate the allocation of resources for higher education institutions, according to pre-assigned parameters. In a word, in a very competitive global educational marketplace, having satisfied "customers" has become an important component of quality assurance and quality enhancement [71]. During the last decades, in the Italian context, there has been both an increased pressure on service quality [68] [59] and a rising interest in evaluating university courses [58]. Since 1999, each Italian public university has begun to monitor and assess the quality of its own teaching activities, with the overall aim of improving the quality of services offered to students [62].

The fundamental regulatory hub that gave rise to the latest changes can undoubtedly be identified in Law no. 240/2010, known as the Gelmini Reform. In order to increase the quality and efficiency of the university system, this rule has profoundly changed the governance structures, funding mechanisms and staff recruitment of the universities. Likewise, with a view to periodic evaluation of the study courses themselves, the ministerial legislation provides for a large set of indicators. In the new set of indicators outlined by Law 240/2010, we can distinguish those related to teaching, internationalization and evaluation of teaching. A continuous and systematic check of the university's performance in the organization, research, and teaching activities is also carried by the Evaluation Unit with the overall goal of the improvement of the internal self-assessment and the promotion of merit. Besides, it should be considered the periodic accreditation, on a three year basis, of study courses. The accreditation decisions are taken consistently with a set of additional requirements, that constitute ANNEX 8 of the AVA 2.2 guidelines, currently in force ([www.anvur.it](http://www.anvur.it)). For these reasons, the assessment of quality within higher education institutions have gained in frequency and proceeds apace. Universities systematically collect learning satisfaction data and measure students' satisfaction in order to monitor and improve the teaching and learning experience [4] and to identify their strengths and areas for enhancement [25, 51, 84]. Student evaluation of teaching is an acknowledged instrument, used to map the quality of teaching in universities and advance the chances that desired learning outcomes will be attained [23, 42, 57].

One of the widely used methods for evaluating courses, programmes and teaching is represented by students' ratings, that are commonplace in many universities. Students' opinions, whose validity has been sufficiently well established (see, among others, [3, 8, 45, 52]), are a vital source of information for both universities and government bodies. Collecting students' evaluations on

the perceived quality of teaching, necessarily requires to take into account the latent construct of the variable of interest. Students' ratings, provided on a set of items and typically expressed on an ordinal scale, are in fact indirect measurement of the quality of the university courses. Thus, instead of assigning them arbitrary scores and analyzing them through linear models, it is desirable to recognize their discrete nature and provide data analysis via probabilistic models [67], such as those of the Item Response Theory (IRT) class. IRT provides a convenient and versatile family of measurement models concerned with the measurement of an individual's latent trait, such as attitude, ability, skill, satisfaction with a product or service, assessed indirectly by a group of items [21, 70]. Their mathematical characteristics allow a transformation from binary or ordinal answer pattern, e.g. Likert type data, into measure on an equal-interval scale [18]. Following this line of research, the focus of the study is to look for patterns of change amid pandemic period in students' evaluation of academic teaching. To this end, a particular class of IRT models, described in the next section, assumes relevance.

### 3 Methodological framework: IRT with relaxed assumptions

IRT encompasses a variety of statistical methods, differing each other in terms of item characteristics that are included in the model and in terms of the response option format [21].

All IRT models include a latent trait parameter and one or more item location parameters, and link the probability of a subject  $i$  ( $i = 1, \dots, N$ ) of providing a certain answer  $Y_{ij}$  to an item  $j$  ( $j = 1, \dots, J$ ) given the individual latent trait and one or more item characteristics.

The basic IRT model for dichotomous data (and a single dimension), also known as the one-parameter Logistic Model [19], or Rasch Model, can be expressed as:

$$Pr(Y_{ij} = 1 | \theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} \quad (1)$$

In the Rasch Model, we have  $J$  parameters  $\beta_j$ , one for each item, and  $N$  parameters  $\theta_i$ , one for each subject. In the original context of application, that is the educational measurement,  $\beta_j$  parameters are a linear continuous measure of "difficulty" and  $\theta_i$  parameters are a linear continuous measure of "ability". An IRT model for dichotomous data operates under the assumption that the items share a common underlying construct (*unidimensionality*), i.e. it is supposed that the item responses are explained by one latent trait, and *local independence*. This last condition implies that, given the value of the latent trait, the item responses are distributed independently.

Motivated by a broad range of applications across different scientific disciplines, considerable advances have been made in the IRT research over the last decades. Particular relevant for the present paper, is the work conducted to using IRT models for measuring change over time. When the objective is

to monitor the trend of a latent variable to understand if significant change occurs, different models, derived from the Rasch models, can be used. Specifically, the Linear Logistic Test Model (LLTM), the Linear Logistic Models with Relaxed Assumptions with the hybrid version (HLLRA) and Mixed Rasch Models (MRM) [26, 29, 73], have been proposed for almost three decades now for the analysis of repeated measure designs, even if not all of them were primarily developed for measuring change.

Here, we restrict our attention to the Linear Logistic models with Relaxed Assumption (LLRA), firstly introduced by [28] whose underlying assumptions are less restrictive than for the other models. LLRA models, in fact, require neither unidimensionality of items nor distributional assumption about the latent trait. Besides, LLRA models provide a flexible framework for modeling change as a function of trend effects, covariate/treatment main and interaction effects [40]. This gives the benefit to test specific hypotheses, for example whether there is a trend effect or whether an experimental group shows more improvement than the control one. Typically, in applications dealing with the measurement of change, the same “real” items are presented several times.

For sake of simplicity, we focus on model for dichotomous response at two time points  $T_1$  and  $T_2$ . A very straightforward paradigm for measuring change is to describe any change of latent variable as a change of item parameters. More in detail, LLRA formalize the above-mentioned effects by a linear decomposition of “virtual items”. An item  $I_j$  presented on two occasions to the same person can be conceived as a pair of “virtual items”  $I_a^*$  and  $I_b^*$ , with associated “virtual” difficulty parameters  $\beta_a^*$  and  $\beta_b^*$ . By assuming that the amount of change between  $T_1$  and  $T_2$  is equal to a constant  $\delta$ , over all subjects, it results that the item parameters are a linear combination of the real item parameter and the change effect. Hence, the pair of “virtual items” generated by the real item  $I_j$  with difficulty  $\beta_j$  is characterized by the two parameters  $\beta_a^* = \beta_j$  and  $\beta_b^* = \beta_j + \delta$ . Let us denote with  $i = 1, \dots, N$  subjects,  $t$  the measurement occasion and  $\theta_{it}$  the “ability parameter” for  $i$ -th person at  $t$  time. A LLRA model is described by the following equations:

$$\Pr(Y_{ij1} = 1 | T_1, \theta_{ij}, \beta_j) = \frac{\exp(\theta_{ij} - \beta_j)}{1 + \exp(\theta_{ij} - \beta_j)} \quad (2)$$

$$\Pr(Y_{ij2} = 1 | T_2, \theta_{ij}^*, \beta_j) = \frac{\exp(\theta_{ij}^* - \beta_j)}{1 + \exp(\theta_{ij}^* - \beta_j)} = \frac{\exp(\theta_{ij} + \delta_{ij} - \beta_j)}{1 + \exp(\theta_{ij} + \delta_{ij} - \beta_j)} \quad (3)$$

where  $\Pr(Y_{ij1} = 1 | T_1, \theta_{ij}, \beta_j)$  and  $\Pr(Y_{ij2} = 1 | T_2, \theta_{ij}^*, \beta_j)$  are the probabilities for subject  $i$  to score 1 on item  $j$  at  $T_1$  and  $T_2$  respectively, while  $\theta_{ij}$  (the latent trait) is the location of subject  $i$  on  $j$ -th item at  $T_1$  and  $\delta_{ij} = \theta_{ij}^* - \theta_{ij}$  is the quantity of change, on item  $j$  for  $i$ -th subject.

The flexibility of LLRA arises from a linear reparametrization of  $\delta_{ij}$  to include different effects. Following [40], the quantity  $\delta_{ij}$ , that is the sum of all changes that subject  $i$  undergoes between two time points, can be modelled as the

sum of group-specific (treatment) effect parameters and an unspecified trend parameter, i.e. changes independent of the treatment, e.g. retest effects. The re-parametrization equation is

$$\delta_{ij} = \sum_s q_{isj} \lambda_{sj} + \tau_j + \sum_{s < l} q_{isj} q_{ilj} \rho_{slj} \quad (4)$$

In Equation 4,  $q_{isj}$  stands for the dosage of treatments  $s$  for item  $j$  in subject  $i$ ,  $\lambda_{sj}$  denotes the effect of the treatment  $s$  on item  $j$ ,  $\tau_j$  is the parameter for the trend effect on item  $j$ , between  $T_1$  and  $T_2$ , and  $\rho_{slj}$  are the parameters for interaction effects of treatments  $s$  and  $l$  on item  $j$  [40]. The interested reader is referred to [27] and [30], where a more accurate description of LLRA models is provided. Here, it is sufficient to say that the parameter estimation in a LLRA model is based on conditional maximum likelihood estimation [48].

## 4 Data

As previously pointed out, this study is aimed at assessing variations in students' feedback on learning experience after the unplanned and rapid interruption of face-to-face education due to COVID-19 pandemic. In response to the COVID 19 pandemic, courses moved to remote instruction on 11st March 2020. Overall, 41 teaching courses, activated at the Department of Philosophical, Pedagogical and Economic-Quantitative Sciences of University of Chieti-Pescara (Italy), have been evaluated. Most of the courses belong to humanistic area (60%), whereas the 40% pertain to the scientific subjects. The number of enrolled students at 41 courses is 2,403. From the university register, we elicited some students characteristics: gender, age, level of study and area of residence. Demographic mix of students population is reported in Table 1. It results that the gender of 70% of students is female and 30% are male. Table 1 shows that the age of more than 50% of them are between 19 and 24 years and about 13% of them are above 35 years old. Almost the number of enrolled students from all study levels are equal. Most of the students (60%) are from Abruzzo and the rest from other regions. Additionally, administrative records contain information about the high schools that students attended throughout their educational careers. We found out that approximately 40% of them attended five years of lyceum while 50% come from technical and professional high schools. Students' satisfaction with some features of university teaching, was assessed quantitatively using an online survey. Face-to-face and online learners of the same courses, taught in two subsequent semesters, before and during the lockdown, were encouraged to fill the questionnaire at the end of regular class times to enhance the learning experience. Respondents were informed that survey is anonymous and out of 2,403 of enrolled students, 1,606 have answered the questionnaire. The survey included questions mainly related to the satisfaction with the course topic, the teaching load, the adequacy of learning material and its content, the respect on schedule with program, the clearness of exams rules, the teacher's clarity and teacher's ability to prompt student's interest in the lecture and the teacher's availability during the lesson

**Table 1** Demographic mix of enrolled students

		N.	%
<i>Gender</i>	Male	731	30
	Female	1672	70
<i>Age</i>	19-22 ys	743	30.9
	22-24 ys	503	20.9
	24-27 ys	449	18.7
	27-35 ys	401	16.7
	over 35 ys	307	12.8
<i>Level of Study</i>	Level 1	822	34.2
	Level 2	829	34.5
	Level 3	752	31.3
<i>Residence</i>	Abruzzo	1480	60
	Other regions	923	40
<i>High School Degree</i>	Lyceum	950	39.5
	Technical Schools	512	21.3
	Professional Schools	145	6
	Other	796	33.1

and/or reception to provide clarifications on the topics covered.

The items were measured on four categories according to the Likert scale: definitely no (1), more no than yes (2), more yes than no (3), definitely yes (4). The questionnaire on the perceived quality of teaching does not allow to gather students' background data. Accordingly, without losing in generality, we can infer representative socio-demographic information of respondents by applying population weights to the survey sample. It is important to stress that the IRT modelling, reviewed in Section 3, and able to track latent trait change over time, is applicable in measurement contexts in which items are administered more than once to the same individuals. In our application, given that it is not possible to rely on the same group of students to acquire opinions about university teaching satisfaction over the two considered measurement occasions (Spring 2020 semester and the semester before), we refer to each teaching course as “subject” in the IRT framework jargon. The starting point for carrying out the analysis was the computation of the overall median of item scores at two time points. To create binary data, responses to the items have been dichotomized according to the overall median value.

## 5 Results

In this section, we present an application of the methodology described in Section 3 to highlight its usefulness in assessing the impact of COVID-19 pandemic on university teaching satisfaction. Quantitative data gathered from the online survey were analyzed using eRm R package [49].

Before assessing the teaching course satisfaction during the transition from traditional face-to-face to online learning, we measure the teaching quality carrying out the dichotomous Rasch model analysis. As known, the essence



of Rasch modelling is to bring the difficulty of item and the ability of subject to the same logit scale. It is worth noting that the meaning of item and person parameters needs to be specified in accordance with the latent trait considered. In our context, borrowing the interpretation from the literature on service quality and customer satisfaction, the item parameter represents the “intrinsic quality” of each particular teaching aspect considered, while the person parameter has an interpretation in terms of satisfaction.

Here, teaching quality is viewed as a unidimensional latent construct, investigated through seven items, listed in Table 2.

Firstly, for a meaningful interpretation of dichotomous Rasch model estimates, we check whether data approximate the key requirements for the model. For testing dimensionality, we relied on the generalized Martin-Löf test [36], that is likelihood test for comparing the null hypothesis of unidimensionality against the alternative hypothesis of two dimensions. An inspection of Martin-Löf test results suggests that the assumption of unidimensionality is tenable on both temporal occasions. Next, to investigate how well data conformed to the Rasch model and identify mis-fitting items, we also consider Rasch fit statistics [77]. In Tables 3 and 4, we display the two commonly used mean square fit statistics, namely the Infit mean square (also referred to as the weighted mean square) and the Outfit (or unweighted) mean square. The Infit and Outfit mean squares can be converted to an approximately normalised t-statistic. These Infit/Outfit t-statistics have an expected value of 0 and a standard deviation of 1 and the results between +2 and -2 suggest that observed data follow the Rasch model while values greater than +2 are interpreted as demonstrating more variation than predicted. As we can see from Table 3, item 4, at time  $T_0$ , is outside the acceptable range of Outfit and Infit statistics. This result indicates that individuals may be responding different than the model predicts and implies a lack of unidimensionality that should be investigated. A graphical representation, useful to discover how persons' distribution relates to the items, is given by the Person-Item map. In an ideal situation, this map should display item parameter and person parameter mirroring each other about a mean of 0 logits. For our data, the Person-Item maps are shown in Figure 1 and they highlight that there are areas of latent construct poorly assessed by the items. A further step in the Rasch analysis is to look at the item parameter estimates, summarized in Table 5. It is important to recall that the smaller values of these estimates describe the “easiest items”, that is features of the teaching with a satisfactory quality level. At both times,  $T_0$  and  $T_1$ , we see that the items to which are associated the highest level of quality are those investigating teacher's availability in providing clarification on the topics (item 7,  $\beta_{T_0} = -2.639$  and  $\beta_{T_1} = -3.182$ ) and teacher's respect or schedule of lessons, exercises, didactic activities (item 3,  $\beta_{T_0} = -0.793$  and  $\beta_{T_1} = -1.628$ ). The Item Characteristics Curves, displayed in Figure 2, allow to compare and identify features with satisfactory and unsatisfactory quality level.

In ascertaining if teaching facets undergone an improvement or a worsening when there was the shift to online distant learning, we have to look at the

**Table 2** Item description

	Item wording
item1	adequacy of teaching material
item2	consistency of teaching with website program
item3	respect for the schedule of lessons, exercises, didactic activities
item4	clearness of examination procedures
item5	teacher's ability in motivating interest in the discipline
item6	teacher's ability in explaining the subject arguments
item7	teacher's availability in providing clarifications on the topics

**Table 3** Item fit statistics at time  $T_0$ 

Item	chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
Item1_T0	20.260	30	0.910	0.654	0.88	-0.33	-0.45
Item2_T0	13.233	30	0.997	0.427	0.573	-1.77	-2.26
Item3_T0	24.122	30	0.766	0.778	0.938	-0.51	-0.19
Item4_T0	68.797	30	0.000	2.219	1.802	2.30	3.08
Item5_T0	16.500	30	0.978	0.532	0.718	-1.33	-1.37
Item6_T0	18.446	30	0.951	0.595	0.741	-1.27	-1.19
Item7_T0	12.743	30	0.998	0.411	0.766	-0.45	-0.77

Fisher's LLRA procedure estimates. To continue this investigation, we need to properly defining the design matrix [40]. Initially, to inspect for change over time, we specify a model where the trend is supposed to be different for all items. In our context, we estimate seven change parameters  $\tau$ . Results summarised in Table 6, suggest that between the two time points all items experienced an improvement in terms of intrinsic quality. In fact, we record significantly positive values for all trend parameters. We also specify a model whose underlying hypothesis is to assume only one trend parameter for all items. For the model, that generalises the trend over all items, the resulting estimate of  $\tau$  is equal to 1.455 with SE=0.242. The two competing models have been compared by performing a likelihood ratio test. The result of test (Likelihood ratio statistic: 7.16, df = 6 p = 0.305) supports the model with just one change parameters for the items.

Our analysis of the effects of COVID-19 pandemic lockdown on academic teaching satisfaction is completed through the specification of a further model where, in addition to the temporal effects, we estimate the influence of certain covariates over time. Specifically, our aim is to assess to what extent the temporal changes may be affected by the grouping of teaching courses in humanistic and scientific classes. To this end, we adopt a LLRA model with a simplification of the re-parametrization of quantity of change. It follows that Equation 4 only includes time (trend) and covariate effects. We assume as "control group" the humanistic class, while the scientific class id the "treatment group". For this latter model, we found significant estimates for item 1 and item 4. Focussing on the teaching courses belonging to the humanistic group, these findings suggest that both the adequacy of teaching material and the clearness of examination procedures exhibit a remarkable improvement when there has been the shift to distant learning.

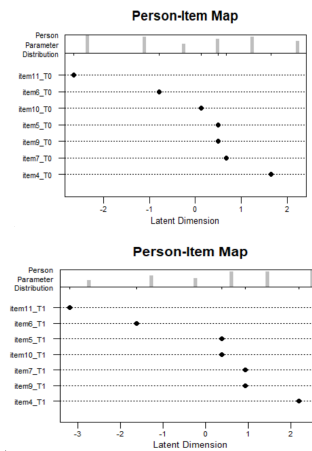
*Looking for patterns of change amid pandemic period in students' evaluation of academic teaching*

**Table 4** Item fit statistics at time  $T_1$

Item	chisq	df	p-value	Outfit MSQ	Infit MSQ	Outfit t	Infit t
Item1_T1	23.097	27	0.680	0.825	1.034	0.00	0.23
Item2_T1	28.718	27	0.375	1.026	1.028	0.19	0.20
Item3_T1	76.595	27	0.000	2.736	1.023	1.67	0.18
Item4_T1	27.494	27	0.437	0.982	1.145	0.09	0.74
Item5_T1	14.254	27	0.979	0.509	0.652	-1.34	-1.87
Item6_T1	11.528	27	0.996	0.412	0.537	-1.93	-2.41
Item7_T1	8.777	27	1.000	0.313	0.874	0.15	-0.02

**Table 5** Item parameter estimates at time point  $T_0$  and  $T_1$

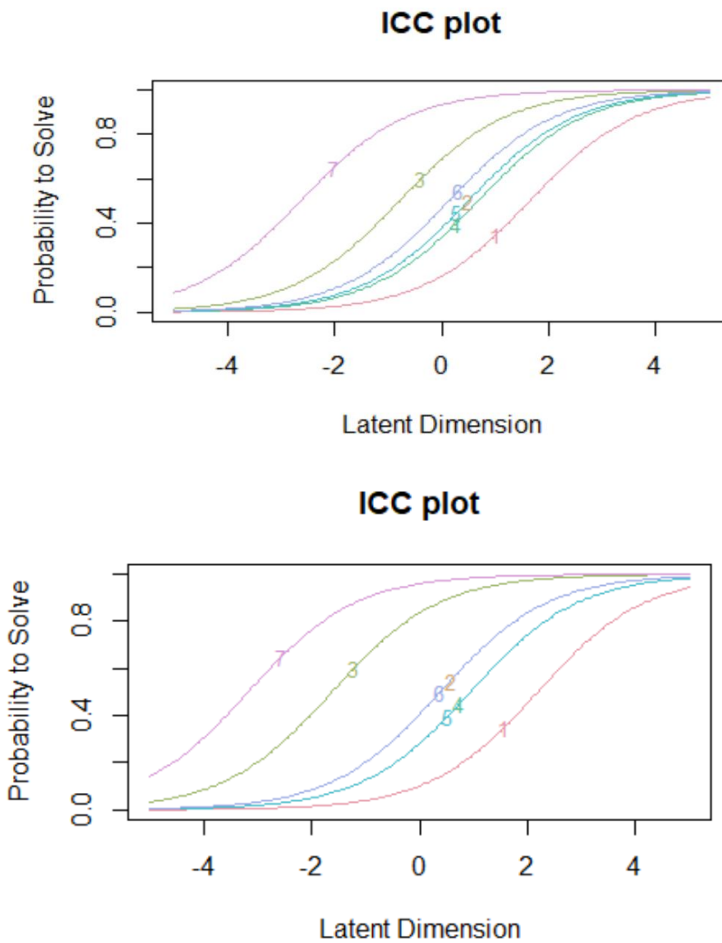
Item	Difficulty parameter at T0	SE	Difficulty parameter at T1	SE
Item1	1.642	0.45	2.184	0.471
Item2	0.493	0.403	0.377	0.441
Item3	-0.793	0.415	-1.628	0.587
Item4	0.675	0.406	0.936	0.433
Item5	0.493	0.403	0.936	0.433
Item6	0.129	0.402	0.377	0.441
Item7	-2.639	0.557	-3.182	0.929



**Fig. 1** Item-person map at time  $T_0$  (left panel) and  $T_1$  (right panel)

## 6 Conclusion and research limitations

Closure and distancing measures caused by the coronavirus outbreak had major impacts on teaching and learning in the university context. Firstly, face-to-face instruction has been converted to online learning. Secondly, lockdowns made households the prime learning space for students. In a nutshell, higher education systems have undergone and will undergo a digital revolution: conference call, online lectures, online exam and communications in virtual environments, will persist in the post-pandemic world. Since e-learning became the unique system for instruction, the concerns are centered on teacher's and learner's experience with this technology and the improvement for the future,



**Fig. 2** Item characteristic curves at time  $T_0$  (top panel) and  $T_1$  (bottom panel)

**Table 6** Trend parameter estimates

Item	$\tau_j$	SE	Lower CI	Upper CI	Sign.
Item1	1.299	0.651	0.023	2.576	***
Item2	1.609	0.632	0.37	2.849	***
Item3	2.047	1.005	0.009	3.026	***
Item4	1.099	0.516	0.086	2.111	***
Item5	1.03	0.521	0.008	2.051	***
Item6	1.253	0.567	0.142	2.364	***
Item7	2.079	1.061	0.004	4.158	***

Sign. codes: \*\*\* 0.05

and on examining whether this method is effective, especially when compared to face-to-face lessons. To this end, in this research, we analysed data collected through a survey administrated on two measurement occasions (before and during COVID-19 pandemic) to obtain accurate opinions on 41 courses taught at University of Chieti-Pescara (Italy). An IRT modelling has been adopted to study response data obtained through a set of items of a questionnaire. Specifically, we relied on IRT models with relaxed assumptions to understand if significant change occurred in the feedback on learning experience after the unplanned and rapid interruption of face-to-face education due to COVID-19 pandemic. Overall, this study has shown that, in this first experience, students had a positive adaptation to online teaching and learning. Our survey findings have in fact demonstrated the satisfaction with the courses or lesson activities delivered in online environment. Positive feedbacks of online teaching experiences are mainly related to instructors' availability for questions and comments and teacher's respect or schedule of lessons, exercises, didactic activities. These results are in line with what has already highlighted in literature, from other researchers in different countries and universities. In this respect, [14] state that the most positive impact with online learning experiences is the class structure that supports flexibility, organization, and clear expectations. Different reasons have led us towards the research carried out in this study. Basically, our first objective is of course related to the necessity of monitoring and improving the teaching. Likewise important is the awareness that COVID-19 pandemic has drastically changed the lives of people all over the world; thus the effectiveness of new teaching mode can be supportive to teaching and learning process post COVID-19. This survey attempted to quickly capture the changes imposed by the emergency precisely to support a community - that of students - potentially very vulnerable in the face of the emergency. Due to their social responsibility roles, universities are obliged to bridge the educational and social challenges of their surrounding societies; the educational systems have to drive the change in order to improve the community readiness for such mode of learning. We believe that in careful designing for learning in higher education post COVID 19 crisis, universities are prompted to adopt different forms of learning activities; in other terms they should creatively merge online and face-to face teaching lessons. Throughout these concluding remarks, we have desired not only to propose a discussion of results but also to stress how this pilot study can be a listening opportunities of students, to make clear the main challenges of this unprecedented education scenario. However, this study is limited since analyzed data are referred to a single department while, for a more comprehensive analysis, it is convenient to enlarge the research to all departments. Additionally, no perspective of universities or faculty members was taken into consideration. The next and deeper research should also investigate the equitable and inclusive access of students to digital learning resources and if their social and emotional needs are being met. With reference to this latter aspect, some researchers show that under the conditions of autonomous learning, motivated students actually improved their academic performance [?

]. Future research should overcome these limitations to come up with a set of useful recommendations that would improve the distance teaching experience.

## Declarations

Not applicable

## References

- [1] Aarrevaara, T., Dobson, I., Elander, C.: Brave new world: Higher education reform in Finland, *HEMP* **21** (2), 1–18 (2009)
- [2] Aristovnik, A., Keržič, D., Ravšelj, D., Tomaževič, N., Umek, L.: Impacts of the COVID-19 pandemic on life of higher education students: A global perspective, *Sustainability* **12** (8438), 1–34 (2020)
- [3] Arubayi, E.: Improvement of instructor and teaching effectiveness: are student ratings reliable and valid?, *High. Educ.* **16** (3), 267–278 (1987)
- [4] Arbaugh, J.B.: System, scholar, or students? Which most influences online MBA course effectiveness?, *J. Comput. Assist. Tomogr.* **30** (4), 349–362 (2014)
- [5] Bertaccini, B., Bacci, S., Petrucci, A.: A graduates' satisfaction index for the evaluation of the university overall quality, *Socio-Econ. Plan. Sci.* **73**, 100875 (2021)
- [6] Boitier, M., Riviere, A.: Freedom and responsibility for French universities: From global steering to local government, *Account. Audit. Account. J.* **26** (4), 616–649 (2013)
- [7] Bolliger, D.U., Halupa, C.: Online student perceptions of engagement, transactional distance, and outcomes, *J. Distance Educ.* **39** (3), 299–316 (2018)
- [8] Cashin, W.E.: Student ratings of teaching: The research revisited. (Idea Paper No. 32). Manhattan: Center for Faculty Evaluation and Development, Kansas State University, (IDEA Paper No. 32). Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- [9] Carini, R., Kuh, G., Klein, S.: Student engagement and student learning: testing the linkages, *Res. High. Educ.* **47** (1), 1–32 (2006)
- [10] Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P., Lam, S.: COVID-19: 20 countries higher education intra-period digital pedagogy responses, *J. Appl. Learn. Teach.* **3** (1), 4–13 (2020)

*Looking for patterns of change amid pandemic period in students' evaluation of academic teaching*

- [11] Chung, E., Subramaniam, G., Dass, L.C.: Online learning readiness among university students in Malaysia amidst (COVID-19), *Asian J. Univ. Educ.* **16** (2), 46–58 (2020)
- [12] Columbu, S., Porcu, M., Sulis, I.: University choice and the attractiveness of the study area: Insights on the differences amongst degree programmes in Italy based on generalised mixed-effect models, *Socio-Econ. Plan. Sci.* **74**, 100926 (2021)
- [13] Contreras, I., Lozano, S.: Allocating additional resources to public universities. A DEA bargaining approach, *Socio-Econ. Plan. Sci.* **71**, 100752 (2020)
- [14] Crews, T., Butterfield, J.: Data for flipped classroom design: Using student feedback to identify the best components from online and face-to-face classes, *High. Educ. Stud.* **4** (3), 38–47 (2014)
- [15] Dal Molin, M., Turri, M., Agasisti, T.: New public management reforms in the Italian universities: Managerial tools, accountability mechanisms or simply compliance?, *Int. J. Public Adm.* **40** (3), 246–269 (2017)
- [16] Deidda Gagliardo, E.: *Il valore pubblico. La nuova frontiera delle performance*, Collana Rirea Aziende (n.30), RIREA, Roma, 2015.
- [17] Dhawan, S.: Online learning: A panacea in the time of COVID-19 crisis, *J. Educ. Technol. Syst.* **49** (1), 5–22 (2020)
- [18] de Ayala, R.J.: The theory and practice of item response theory, *Psychometrica* **75**, 778–779 (2009)
- [19] De Boeck, P. Wilson, M.: *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, Springer-Verlag, New York, 2004.
- [20] Del Bene, L.: *Lineamenti di pianificazione e controllo per le amministrazioni pubbliche*, Giappichelli, Torino, 2008.
- [21] De Mars, C.: *Item response theory. Understanding statistics measurement*, Oxford University Press, 2010.
- [22] Donna, G.: *L'Università che crea valore pubblico*, Il Mulino, Bologna, 2018.
- [23] Emerson, J., Mosteller, F., Youtz, C.: Students can help improve college teaching: a review and an agenda for the statistics profession, in: Rao CR, Székely GJ eds. *Statistics for the 21st century: methodologies for applications of the future*, 2000.

- [24] Engwall, L.: Universities, the state and the market: Changing patterns of university governance in Sweden and beyond, *J. High. Educ. Policy Manag.* **19** (3), 87–103 (2007)
- [25] Eom, S.B., Wen, H.J., Ashill, N.: The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation, *Decis. Sci. J. Innov. Educ.* **4** (2), 215–235 (2006)
- [26] Fischer, G.H.: The linear logistic test model as an instrument in educational research, *Acta Psychol.* **37**, 359–374 (1973)
- [27] Fischer, G.H.: Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen, H. Huber., 1974.
- [28] Fischer, G.H.: Some probabilistic models for measuring change, in: In D N M de Gruijter & L J Th van der Kamp (Eds.), *Advances in psychological and educational measurement*, New York: Wiley, 1976.
- [29] Fischer, G.H.: Logistic latent trait models with linear constraints, *Psychometrika* **48**, 3–26 (1983)
- [30] Fischer, G.H., Ponocny, I.: Extended rating scale and partial credit models for assessing change, in: *Rasch models. Foundations, recent developments and applications*, Springer-Verlag, New York, 1995, pp. 353–370.
- [31] Flores, M.A., Brown, G., Pereira, D., Coutinho, C., Santos, P., Pinheiro, C.: Portuguese university students' conceptions of assessment: taking responsibility for achievement, *High. Educ.* **79**, 377–394 (2020)
- [32] Flores, M.A., Barros, A., Simão, A.M.V., Pereira, D., Flores, P., Fernandes, E., Costa, L., Ferreira, P.C.N.: Portuguese higher education students' adaptation to online teaching and learning in times of the COVID-19 pandemic: personal and contextual factors, *High.Educ.* (2021)
- [33] Fredricks, J.A., Filsecker, M., Lawson, M. A.: Student engagement, context, and adjustment: addressing definitional, measurement, and methodological issues, *Learn Instr.* **43**, 1–4 (2016)
- [34] Gonzalez-Gomez, D., Jeong, J.S., Rodriguez, D.A.: Performance and perception in the flipped learning model: An initial approach to evaluate the effectiveness of a new teaching methodology in a general science classroom, *J. Sci. Educ. Technol.* **25** (3), 450–459 (2016)
- [35] Gordon, N., Grey, S., Brayshaw, M.: *Motivating and engaging students through technology*, Student Engagement , New York: Nova Science Publishers Inc., J. Hawkins (2015) 25–43



*Looking for patterns of change amid pandemic period in students' evaluation of academic te*

- [36] Glas, C.A.W., Verhelst, N.D.: Testing the rasch model, in: In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications*, New York: Springer-Verlag, 1995, pp. 69–95.
- [37] Gray,J., DiLorento, M.: The effects of student engagement, student satisfaction, and perceived learning in online learning environments, *Int. J. Leadersh. Educ.* . **11** (1), 98–119 (2016)
- [38] Guangul, F.M., Suhail, A.H., Khalit, M.I., Khidhir, B.A.: Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College, *Educ. Assess. Evaluation Account.* **32**, 519–535 (2020)
- [39] Harley,J., Lou,N., Liu, Y., Cutumisu, M., Daniels, L., Leighton,J., Nadon, L.: University students' negative emotions in a computer-based examination: the roles of trait test-emotion, prior test-taking methods and gender, *Assess. Eval. High. Educ.* **46** (6) (2021)
- [40] Hatzinger, R., Rusch, T.: IRT models with relaxed assumptions in eRm: A manual-like instruction, *Psychol. Sci. Q.* **51**, 87–120 (2009)
- [41] Hodges, C., Moore, S., Lockee, B., Trust, T., Bond, A.: The Difference Between Emergency Remote Teaching and Online Learning, <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning> (2020)
- [42] Kember, D., Ginns, P.: *Evaluating teaching and learning*, Routledge, 2012.
- [43] Korkmaz, G., Toraman, C.: Are we ready for the (post-COVID-19) educational practice? An investigation into what educators think as to online learning, *Int. J. Sci. Educ.* **4** (4), 293–309 (2020)
- [44] Krejsler, J.: Discursive battles about the meaning of university: The case of Danish university reform and its academics, *Eur. Educ. Res. J.* **5** (3-4) 210–220 (2006)
- [45] Kulik, J.: Student ratings: validity, utility and controversy, *New dir. inst. res.* **27**, 9–25 (2011)
- [46] Liguori, E.W., Winkler,C.: From Offline to Online: Challenges and Opportunities for Entrepreneurship Education Following the COVID-19 Pandemic, *J. Entrep. Educ.* **3** (4), 346–351 (2020)
- [47] Lockman, A.S., Schirmer, B.R.: Promising, research based, and evidence-based practices, *J. Educ. e-Learn. Res.* **7** (2), 130–152 (2020)

- [48] Mair, P., Hatzinger, R.: CML based estimation of extended Rasch models with the eRm package in R, *Psychol. Sci.* **49** (1), 26–43 (2007)
- [49] Mair, P., Hatzinger, R.: Extended Rasch Modeling : The eRm Package for the Application of IRT Models in R, *J. Stat. Softw.* **20** (9), 1–20 (2007)
- [50] Manes Rossi, F., Nicolò, G., Tartaglia Polcini, P.: New trends in intellectual capital reporting: Exploring online intellectual capital disclosure in italian universities, *J. Intellect. Cap.* **19** (4), 814–835 (2018)
- [51] Marsh, H. W.: Seeq: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching, *Br. J. Educ. Psychol.* **52** (1), 77–95 (1982)
- [52] Marsh, H.W: Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility, *Br. J. Educ. Psychol.* **76**, 707–754 (1984)
- [53] Martin, F., Bolliger, D. H.: Engagement matters: student perceptions on the importance of engagement strategies in the online learning environment, *Online Learn. J.* **22** (1), 205–222 (2018)
- [54] Metcalfe, A.: Visualizing the Covid-19 pandemic response in Canadian higher education: An extended photo essay. *Studies in Higher Education, Stud. High. Educ.* **46** (1), 5–18 (2021)
- [55] Mussari, R., D'Alessio, L., Sostero, L.: Il nuovo sistema contabile delle università, *Azienda Pubblica* **3**, 227–246 (2015)
- [56] Naude, P., Ivy, J.: The marketing strategies of universities in the United Kingdom, *Int. J. Educ. Manag.* **13** (3), 126–136 (1999)
- [57] Onwuegbuzie, A.J., Witcher, A.E., Collins, K.M.T., Filer, J.D., Wiedmaier, C.D., Moore, C.: Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis, *Am. Edu. Res. J.* **44**(1), 113-160 (2007)
- [58] Pagani, L., Seghieri, C.: Statistical analysis of teaching effectiveness from students' point of view, in: In: Mrvar A, FerligojA (eds) *Developments in Statistics. Metodološki zvezki* 17, Ljubljana, FDV, Vol. 17, Andrej Mrvar and Anuška Ferligoj (Editors), 2002, pp. 197–208.
- [59] Paletta, A.: *Il governo dell'università. Tra competizione e accountability*, Il Mulino, Bologna, 2005.
- [60] Palihawadana, G.H.: Modeling module evolution in marketing education, *Qual. Assur. Educ.* **7** (1), 41–46 (1999)

*Looking for patterns of change amid pandemic period in students' evaluation of academic te*

- [61] , Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., Sindhi, S.: Online education: Worldwide status, challenges, trends, and implications, *J. Glob. Inf. Manag.* **21** (4), 233–241 (2018)
- [62] Paolini, A., Del Bene, L.: *Monitorare la performance delle università statali per generare valore pubblico*, FrancoAngeli, Milano (2021)
- [63] Pereira, D., Niklasson, L., Flores, M.A.: Students' perceptions of assessment: A comparative analysis between Portugal and Sweden, *High. Educ.* **73**, 153–173 (2017)
- [64] Patterson, G.: The applicability of institutional goals to the university organization, *J. High. Educ. Policy Manag.* **23** (2), 159–169 (2001)
- [65] Pei, L., Wu, H.: Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis, *Med. Educ. Online* **24** (1), 1666538 (2019)
- [66] Rajabalee, Y.B., Santally, M.I.: Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy, *Educ. Inf. Technol.* **26**, 2623–2656 (2020)
- [67] Rampichini, C., Grilli, L., Petrucci, A.: Analysis of university course evaluations: from descriptive measures to multilevel models, *Stat Methods Appt*, **13**, 357–373 (2004)
- [68] Rebora, G.: La valutazione degli Atenei: alla ricerca della qualità, *Azienda Pubblica* **5-6**, 411–437 (2003)
- [69] Rebora, G., Turri, M.: La governance del sistema universitario in Italia: (1989-2008), *Azienda Pubblica* **22** (2), 259–284 (2009)
- [70] Reckase, M.D.: *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*, New York, NY Springer, 2009.
- [71] Rienties, B.: Understanding academics' resistance towards (online) student evaluation, *Assess. Eval. High. Educ.* **39** (8), 987–1001 (2014)
- [72] Saeed, S.: COVID-19 has exacerbated inequality in higher education, <https://www.universityworldnews.com/post.php?story=20201023103200335> (2020)
- [73] Scheiblechner, H.: Das Lernen und Lösen komplexer Denkaufgaben. [The learning and solving of complex reasoning items], *Z. Exp. Angew. Psychol.* **3**, 456–506 (1972)
- [74] Selwyn, N.: *Digital technology and the contemporary university: Degrees of digitization*, Routledge, London, 2014

- [75] Shahzad, A., Hassan, R., Aremu, A.Y, Hussain, A., Lodhi, R.: Effects of Covid-19 in E-learning on higher education institution students: The group comparison between male and female, *Qual. Quant.* **55**, 805–826 (2021)
- [76] Smalley, A.: Higher education responses to coronavirus (COVID-19), <https://www.ncsl.org/research/education/higher-education-responses-to-coronavirus-covid-19.aspx>(2021)
- [77] Smith, R.M, Suh, K.K.: Rasch fit statistics as a test of the invariance of item parameter estimates, *J. Appl. Meas.* **4** (2), 153–163 (2003)
- [78] Stevens, T., Switzer, C.: Differences between Online and Traditional Students: A Study of Motivational Orientation, Self-Efficacy, and Attitudes, *Turk. online J. Distance Educ.* **7** (2), 90–100 (2006)
- [79] Tasso, A., Sahin, N. San Roman, G.: COVID-19 disruption on college students: Academic and socioemotional implications, *Psychol Trauma.* **13** (1), 9–15 (2021)
- [80] Vicente, P.N., Lucas, M., Carlos, V., Bem-Haja, P.: Higher education in a material world: Constraints to digital innovation in Portuguese universities and polytechnic institutes, *Educ. Inf Technol.* **25**, 5815–5833 (2020)
- [81] Wang, C., Xie, A., Wang, W., Wu,H.: Association between medical students' prior experiences and perceptions of formal online education developed in response to Covid-19: A cross-sectional study in China, *BMJ Open.* **10** (10) (2020)
- [82] Watermeyer, R., Crick, T., Knight, C., Goodall, J.: COVID-19 and digital disruption in UK universities: afflictions and affordances of emergency online migration, *High.Educ.* **81**, 623–641(2021)
- [83] Xiong, W., Jiang, J., Mok, K.H.: Kong university students' online learning experiences under the (COVID-19) pandemic, <https://www.hepi.ac.uk/2020/08/03/hong-kong-university-students-online-learning-experiences-under-the-covid-19-pander>(2020)
- [84] Zerihun, Z., Beishuizen, J., Os, W.: Student learning experience as indicator of teaching quality, *Educ. Assess. Evaluation Account.* **24** (2), 99–111 (2012)