MOX-Report No. 69/2025

Measuring Academic Stress and Well-Being in Higher Education: A Psychometric Study

Marino, F.; Guagliardi, O.; Di Stazio, F.; Mazza, E.; Paganoni, A.M.; Tanelli, M.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

Journal Title
XX(X):1–11
©The Author(s) 2016
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Measuring Academic Stress and Well-Being in Higher Education: A Psychometric Study

Filippo Marino[1], Oriana Guagliardi[2], Filomena Di Sanzio[3], Elena Mazza[2,4], Anna Maria Paganoni[1], Mara Tanelli[2]

Abstract

This study investigates the assessment of perceived academic stress and its impact on students' mental health by employing advanced psychometric and statistical models. A dataset of 9,000 university students was analyzed using Item Response Theory (IRT), specifically the Four-parameter Nested Logistic Regression Model (4PLnRM), to estimate latent parameters that quantify well-being across nine stress-related domains. Linear and mixed-effects models were applied to identify the most relevant socio-demographic, psychosocial and academic risk factors, highlighting the strong influence of economic conditions, exam backlog, and academic self-perceptions. To further validate these measures, Random Forest classification models were trained to identify students at different levels of psychological vulnerability psychological risk, demonstrating that latent parameters are effective predictors of distress and well-being.distress. Additionally, Exploratory Factor Analysis (EFA) on self-reported mental health symptoms psychological symptoms revealed four interpretable latent factors—anxiety, depression, motivational block, and somatization—used in subsequent clustering to classify students into low, medium, and high-risk groups. Across methods, consistent associations emerged between risk classes and demographic variables such as gender, age, academic performance, and economic satisfaction. The results emphasize the value of latent psychometric modeling for identifying stress mechanisms and developing targeted interventions aimed at improving the academic climate and supporting students' mental health.

Keywords

Mental health, Psychological disturbs, Item Response Theory Models, latent parameters, Classification, Random Forest, Exploratory Factoral Analysis, Risc Factors, Cluster Analysis

Introduction

The assessment of mental health has long represented a central theme in psychological and social science research, yet developing objective tools that can reliably capture the psychological and physical well-being of individuals and groups remains a complex challenge. In this context, Item Response Theory (IRT) models constitute a widely used approach for analyzing response patterns derived from Likert-scale questionnaires, as they allow researchers to capture the latent nuances that characterize subjective perception.

In this study, we propose the use of the Four-parameter Nested Logistic Regression Model (4PLnRM), an advanced extension of the IRT family, with the aim of more accurately modeling the complexity of data collected through a questionnaire on perceived academic stress, administered at the Politecnico di Milano. The analysis pursues two main objectives: first, to estimate latent parameters that quantify students' well-being; and second, to identify the key academic stressors influencing their university experience.

The results are then integrated into a classification model based on Random Forest, designed to detect student subgroups at increased risk. In this way, the work goes beyond a purely methodological evaluation and offers an applied perspective, highlighting how specific parameters can contribute both to understanding stress mechanisms and to developing targeted interventions to improve the academic climate and support students' mental health.

Finally, in the last section an analysis will be conducted using Exploratory Factor Analysis (EFA) models, based on a subset of the dataset that investigates the presence of certain psychological symptoms among students. The factor scores from the model will be used to perform a clustering of students and categorize them into risk classes, in order to investigate once again whether there are demographic differences and variations in the potential risk categories under study.

[1]MOX Laboratory, Department of Mathematics (DMAT), Politecnico di Milano, Milan, Italy.
[2]Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Milan, Italy.
[3] Campus Life, Politecnico di Milano, Milan, Italy.
[4] Psychiatry Clinical Psychobiology, Division of Neuroscience, IRCCS Ospedale San Raffaele, Milan, Italy.

Corresponding author:
MAnna Maria Paganoni, MOX Laboratory, Department of Mathematics (DMAT), Politecnico di Milano, Milan, Italy
Email: anna.paganoni@polimi.it

Dataset presentation

The dataset we aim to create is inspired by the study conducted in Bedewy and Gabriel (2015). This study provides a comprehensive framework for understanding the various sources of academic stress among university students, which is crucial for developing a dataset that accurately captures the factors contributing to stress in academic environments.

The primary motivation behind creating this dataset is to build upon the findings of Bedewy and Gabriel (2015), who identified key factors that contribute to academic stress among university students. Their study developed the Perception of Academic Stress Scale (PAS), which categorizes stress into main factors:

1. **Pressures related to academic expectancies**: This includes stress related to high expectations from teachers and parents, as well as competitive peer pressure.
2. **Perceptions of workload**: This factor encompasses stress due to excessive academic workload
3. **Academic Self-Perceptions**: This involves students' confidence in their academic abilities, their future career prospects, and their ability to make academic decisions.
4. **Academic satisfaction**: This factor relates to stress caused by limited time for classes, homework, and relaxation.
5. **Life quality and personal well-being**:These are very important and sensible questions related to the personal perception of the quality of life
6. **Social support and personal relationship**:this factor relates to how big is the support from family, students and professors.
7. **University climate**:This factor encompasses the questions related to the university climate and the competition within it.
8. **Motivation towards study**: This factor investigates how motivated an individual is to study, and to what extent they value academic challenge and pressure.
9. **Stress, psychological disturb, academic difficulties**: these are the most sensible questions and this section of the dataset will be used later as output.

These factors were identified as the most significant contributors to academic stress, and they provide a solid foundation for structuring the dataset. By organizing the dataset around these macro-categories, we can ensure that it captures the most relevant and impactful sources of stress, as identified by Bedewy and Gabriel (2015).

Moreover, the study highlights the importance of considering both academic and non-academic factors when assessing stress among students. Indeed, consistent with previous research by Brand and Schoonheim-Klein (2009), stress is described as a multifactorial construct shaped by socio-cultural, environmental, and psychological factors.

The datset was designed by Dr. Elena Mazza from the Polipsi area of the Polytechnic University of Milan. After thorough cleaning, the dataset consists of 9000 rows and 243 columns and they are all categorical data, with a lot of Likert-type scales and demographic/potential risk factors. As for the Likert scales, the response value 5 will always correspond to the most positive possible answer (e.g., 'very good'), while 1 will represent the worst. In this way, the value assigned to the highest response category is standardized. This initial data collection paints a concerning picture regarding the mental health of students, highlighting the need for a deeper understanding of the factors that most significantly contribute to stress. Consequently, identifying these factors and proposing interventions becomes crucial for improving the overall university experience (considering that a state of well-being also positively impacts academic performance). Below are some key findings and examples of questions of the questionary:

- I am satisfied with my degree program.
- In many respects, my life is close to my ideal.
- I am satisfied with my life.
- I am proud to be a student of this university.
- I enjoy challenges.
- I am confident that I will succeed in my future career.

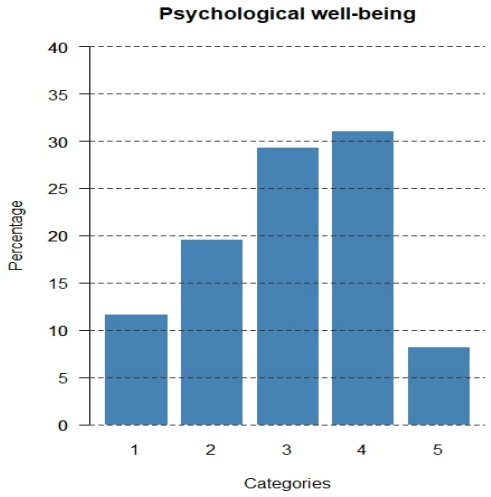

Figure 1. Histogram of Psychological well-being question

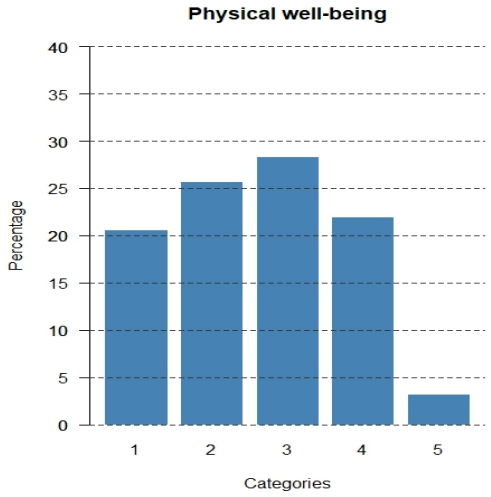

Figure 2. Histogram of Physical well-being question

Item Response Theory Models

The central feature of Item Response Theory (IRT) is the specification of a mathematical function relating the probability of an examinee's response on a test item to an underlying ability. In the 1940s and 1950s, the emphasis was on the normal-ogive response function, but for statistical and practical reasons this function was replaced by the logistic response function in the late 1950s. Today, IRT models using logistic response functions dominate the measurement field as (Van der Linden and Hambleton 1997) said in their Handbook of Modern Item Response Theory. Long experience with tools like thermometers or speedometers may suggest that measurement is a direct numerical reading from a device. However, this is not the case for educational and psychological (Lord and Novick (1968)). Our dataset contains a lot of columns, so a lot of information about each person. The objective of this chapter is to create a score capable of best capturing the general psycho-physical state of the subject.

IRT: key concepts

In psychometrics, the columns of a dataset corresponding to the questions in a questionnaire are usually referred to as *items*. Parametric IRT models are based on two main concepts:

- the latent parameter θ_i, associated with each subject who completes the questionnaire;
- the item difficulty parameter a_i, where $i = 1, \ldots, n$ indexes the items.

The latent parameter θ was originally conceived as a measure of a subject's ability (Rasch (1960)) to correctly answer items in a multiple-choice test. In the context of psychometric questionnaires, however, θ can be interpreted more broadly. For example, it may reflect how happy or sad a person is, depending on the type of items considered (Bech *et al.* (1992),Cressie and Holland (1983)). In our case, a subject with a high θ value can be interpreted as an indicator of better overall psycho-physical well-being. Practically speaking, a higher θ corresponds to a greater probability of selecting higher values on the rating scale (e.g., choosing "5" to indicate a very positive state).

The item parameter a_i, on the other hand, represents the difficulty of an item. In ability tests, a higher value means a more difficult question. In psychometric contexts, where there is no longer a "correct" or "incorrect" answer, this interpretation is less direct. Here, a_i can be understood as how demanding it is for subjects to endorse higher response categories on item i.

Since our goal is to develop a reliable measure of stress, we must select the model that best fits the data and then estimate the latent parameter θ, which will serve precisely as our stress quantifier. There are many ways to construct such a score, but IRT models capture information that, for example, a simple average of responses cannot, thanks to their more sophisticated treatment of the problem.

Four Parameter Nested Logistic Regression Model

The model used in this study will be the $Four\ Parameter\ Nested\ Logistic\ regression\ Model$ (4PLnRM). The reasons for using this model are its performances mainly explained by the factors that other simpler models did not consider (Chalmers (2012),Storme *et al.* (2019), Suh and Bolt (2010)). In order to understand how this model adapts to our data we first need to introduce the $Four\ parameter\ logistic\ regression\ model$, suited for dichotomous data.

Four Parameter Logistic Regression Model

The 4PL model is an extension of the 3-Parameter Logistic regression model (3PL); both these models are used for dichotomous (so 0 and 1) kind of data for ability tests, where one category corresponds to the correct answer to the item i.

In the 3PL, probability of responding correctly (let us say 1) for item i by a subject with ability θ is given by:

$$P(X = 1 \mid \theta) = g + (1 - g)\frac{1}{1 + \exp\left(-a(\theta - d)\right)}$$

where the lower asymptote g (the guessing parameter) represents the probability that an examinee with extremely low ability will correctly answer an item with difficulty a (Harvey and Hammer, 1999).

But the 3PL model might severely penalize a high-ability student who makes a careless error on an easy item (Barton and Lord 1981; Rulison and Loken 2009). More specifically, in the 3PL IRT model the upper asymptote of 1 assigns a probability of 0 to a high-ability student failing to answer an easy item correctly (Loken and Rulison 2010).

Barton and Lord (1981) (1981) introduced an upper asymptote parameter u, which represents an upper bound for $P(\theta)$, with the respect to very high values of θ.

In fact, for each subject j, we now express the probability of responding 1 to item i as:

$$P(X = 1 \mid \theta) = g + (u - g) \cdot \frac{1}{1 + \exp\left(-a(\theta - d)\right)}$$

The IRT models for dichotomous data are structured so that a highly skilled subject, that is, with a very high θ, answers the item correctly ($X = 1$) with probability

$$P(X = 1 \mid \theta) \approx 1$$

Conversely, a subject with a very low θ has probability

$$P(X = 1 \mid \theta) \approx 0$$

However, now we consider the possibility that a subject with a very high or low θ may respond differently than expected to the item.

Specifically:

- For $\theta \to \infty$, $P(X = 1 \mid \theta) \to u$ and no longer 1.
- For $\theta \to -\infty$, $P(X = 1 \mid \theta) \to g$ and no longer 0.

How it works

In the four-parameter nested logistic regression model (4PLnRM) with a nominal component, the probability of selecting a given response category is divided into two cases. First, let us denote the fifth category ($h = 5$) as the key category, which in the context of a Likert-type scale can be interpreted as the highest level of agreement ("strongly agree"). The probability of endorsing this key category follows a four-parameter logistic (4PL) curve:

$$P(X = 5 \mid \theta) \;=\; g + (u - g)\,\frac{1}{1 + \exp[-a(\theta - d)]},$$

where a is the discrimination parameter, d is the location (or difficulty) parameter, g is the lower asymptote, and u is the upper asymptote. This component governs the likelihood that an individual fully endorses the highest category as a function of the latent trait θ.

If the response does not fall into the key category, that is, if $h \in \{1, 2, 3, 4\}$, then the probability is modeled through a nominal model. In this case, the overall probability of selecting a distractor category is weighted by the complement of the 4PL trace line, i.e. $1 - P(X = 5 \mid \theta)$, and then multiplied by the nominal model probability for that category:

$$P(X = h \mid \theta) = [1 - P(X = 5 \mid \theta)] \cdot P^{\text{nominal}}(X = h \mid \theta), \quad h \in \{1, 2, 3, 4\}. \tag{1}$$

The nominal probability is given by

$$P^{\text{nominal}}(X = h \mid \theta) = \frac{\exp(ak_{m-1}(a\theta + \delta_{hk}))}{\sum_{m=0}^{H-1} \exp(ak_m(a\theta + \delta_{mk}))}, \quad h \in \{1, 2, 3, 4\}. \tag{2}$$

where δ_{hk} are the step parameters of the nominal model and $H - 1$ is the number of non-key categories (in this case, $H - 1 = 4$). This structure ensures that the choice among the distractor categories is modeled according to an item response framework appropriate for polytomous data.

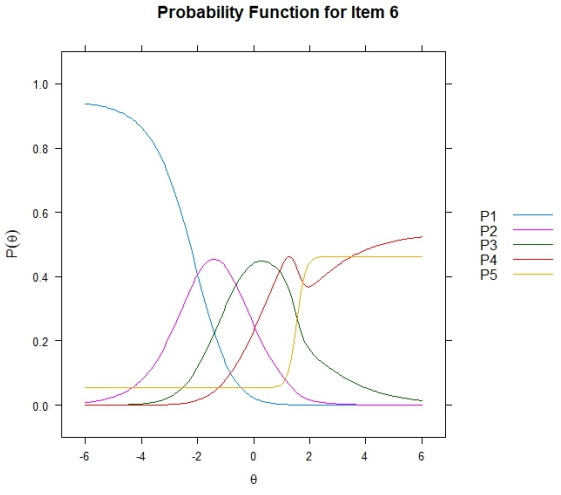

Figure 3. Example of probability curves with 4PL

In summary, the nested model operates in two stages: first, a 4PL function determines the probability of endorsing the highest category; second, conditional on not endorsing this category, the nominal model specifies how the probability mass is distributed across the remaining categories. This allows us to assume that a subject with a high θ can answer a category different than 5 with a probability different from 0; this allows for a more precise estimation of the latent parameter because, in the psychometric context, nothing guarantees that a subject with a high value of the latent parameter will necessarily choose response category 5 for all the questionnaire items, since, for example, they might not feel comfortable in certain areas of the urban environment.

Application of the 4PLnRM model

Previously, the division of the dataset into stress subcategories was introduced, for each of which specific questionnaire items were selected. For each of these divisions, the 4PL model was applied separately in order to obtain from each the latent parameter; these parameters will serve as the variables indicating how 'happy' or not a subject is within each subcategory. In practice, an individual with a high value of theta in the category 'Academic self-perception' will tend to experience academic life in a positive way.

The reason for applying the models separately rather than to the dataset as a whole is that the goal is to investigate how different domains—such as academic aspects, family sphere, or urban environment—affect the overall health of students. In fact, several questions in the questionnaire aim to investigate, in a more general way, how a subject feels; these belong to the subcategory 'Stress, psychological discomfort, academic difficulties'. The latent parameter estimated in this subcategory will be the 'general' quantifier of individuals' happiness, on which further studies will be conducted to identify the main risk factors that influence it.

The result is therefore a new dataset with columns corresponding to the 9 latent parameters estimated by the 4PL model for each student, each associated with one of the 9 subcategories identified in the questionnaire.

Linear Models: which are the most relevant stress factors?

The choice to use linear models (LM) to explain the general well-being of the students was made with the aim of having models that are simple to understand but, at the same time, provide significant results. Subsequently, Linear mixed models (LMM) will be employed to identify which grouping factors (such as gender, age, work-related stress, etc.) have the greatest influence on the student stress.

The output of the model corresponds to the latent parameter associated with the stress subcategories: *Psychological distress*, *Academic difficulties (category number 5)*, and *Life quality and personal well-being (category number 9)*. Specifically, this parameter was estimated by merging the two subcategories that included the most general questions on personal mental health and perceived stress. Therefore, the dependent variable of the linear model can be regarded as the most appropriate measure of individuals' overall stress levels.

To better understand how linear models will be applied, it is first necessary to clarify the variables included in the dataset to be used.

Table 1. Categorical variables of interest

Variable	Description
country	Two levels: Italy, foreign student
home_condition	Three levels: commuter, out-of-town, local student
age	Range of age of belonging
sex	Sex of the subject
degree_kind_triennale	Three levels: ING (engineering), DES (design), ARC (architecture) for bachelor's degree
average_exams	Range of exams' averages
exams_condition	Specifies if a student is up-to-date with exams (5 levels)
economic_stress	Level 1: good economic condition, 0 otherwise

Table 2. Latent parameters estimated by the 4PL model

Category	Latent parameter
1: Pressures related to academic expectancies	$\theta_{\text{academic-expect}}$
2: Perceptions of workload	θ_{workload}
3: Academic Self-Perceptions	$\theta_{\text{academic-perc}}$
4: Academic satisfaction	$\theta_{\text{satisfaction}}$
5: Life quality and personal well-being	$\theta_{\text{wellbeing}}$
6: Social support and personal relationship	$\theta_{\text{social-support}}$
7: University climate	θ_{climate}
8: Motivation towards study	$\theta_{\text{motivation}}$
9: Stress, psychological disturb, academic difficulties	θ_{stress}

Best models and results

Below, we present the models with the variables that proved to be the most relevant with respect to the designated output, $\theta_{\text{stress-wellbe}}$.

- Linear Model

$$\begin{aligned}\theta_{\text{stress-wellbe}} = \beta_0 + \beta_1\,\theta_{\text{academic\_perception}} + \beta_2\,\theta_{\text{academic\_expect}} \\ + \beta_3\,\theta_{\text{workload}} + \beta_4\,\theta_{\text{social\_support}} + \beta_5\,\theta_{\text{satisfaction}} \\ + \beta_6\,\theta_{\text{motivation}} + \beta_7\,\alpha_{\text{economic\_condition}} + \varepsilon_i\end{aligned} \tag{3}$$

$$\varepsilon_i \sim N(0, \sigma^2) \tag{4}$$

$$R^2_{\text{adjusted}} = 0.632 \tag{5}$$

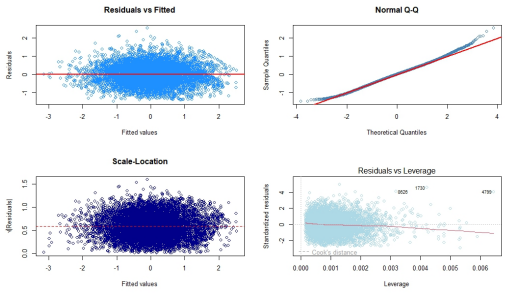

Figure 4. Random intercept plot Mod 3

	Estimate	t value	p-val
(Intercept)	-0.166	-9.392	$< 2e{-}16^{***}$
$\theta_{\text{academic\_perception}}$	0.322	35.313	$< 2e{-}16^{***}$
$\theta_{\text{academic\_expect}}$	0.142	15.499	$< 2e{-}16^{***}$
θ_{workload}	0.172	21.550	$< 2e{-}16^{***}$
$\theta_{\text{social\_support}}$	0.166	21.177	$< 2e{-}16^{***}$
$\theta_{\text{satisfaction}}$	0.092	10.744	$< 2e{-}16^{***}$
$\theta_{\text{motivation}}$	0.259	26.489	$< 2e{-}16^{***}$
$\theta_{\text{economic\_condition}}$	0.215	11.314	$< 2e{-}16^{***}$

Table 3. Regression meaningful results

As we can observe, very high values of R^2 are obtained, indicating that the macro-categories significantly influence the overall stress score as we expected from the study of Bedewy and Gabriel (2015). Additionally, the residuals are homoschedastic. As expected, these stress factors correlate positively with the score and are all statistically significant, although those related to the environment and university climate have a slightly less significant impact. Moreover the economic self-perception factor is very impactfull on the output of the model.

- Linear Mixed Models with random intercept.

Some of the possible risk variables or demographic data had several categories to be taken into account; the goal is to verify whether, in some cases, a pattern emerges in the value of the random intercept when adding such variables to the mixed linear model. The model included all latent parameters of the categories that proved to be relevant in the previous linear model, so that the variables presented below, even if they contribute only a small proportion of explained variability (PVRE), are still relevant for the purposes of the study.

Linear Mixed model with *native country* as a grouping factor

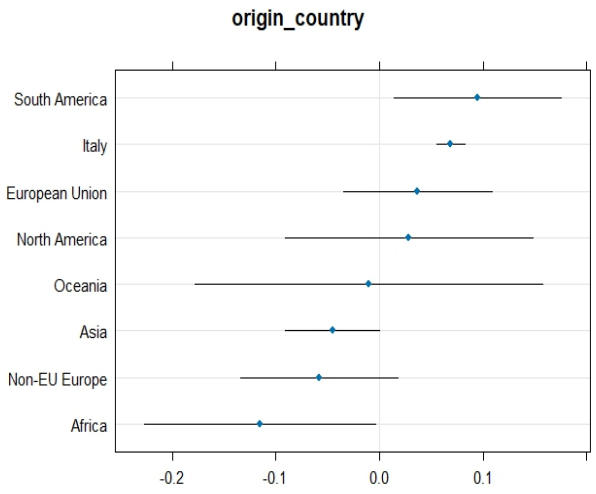

Figure 5. Random intercept plot country of origin

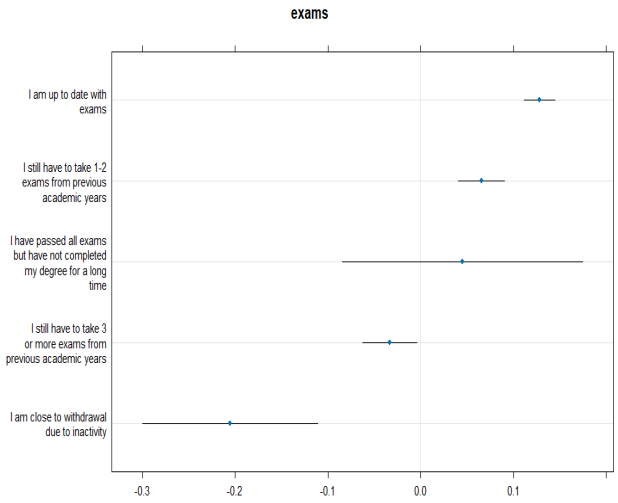

Figure 6. random intercept plot exams condition

$$
\begin{aligned}
\theta_{\text{stress-wellbe}} &= \beta_1\,\theta_{\text{academic\_perception}} + \beta_2\,\theta_{\text{academic\_expect}} \\
&\quad + \beta_3\,\theta_{\text{workload}} + \beta_4\,\theta_{\text{social\_support}} \\
&\quad + \beta_5\,\theta_{\text{satisfaction}} + \beta_6\,\theta_{\text{motivation}} \\
&\quad + b_{\text{country},0j} + \epsilon_{ij} \\
b_{\text{country},0j} &\sim N(0, \sigma_b^2), \\
\varepsilon_{ij} &\sim N(0, \sigma^2).
\end{aligned} \tag{6}
$$

```
PVRE = 0.018
```

Although the PVRE index is not particularly high, differences emerge across categories of the grouping factor: Italian students generally achieve higher scores, while Asian students appear to face greater difficulties. The reasons for this trend cannot be determined with certainty, but it nonetheless represents a relevant factor in our analysis. South American students show a positive random intercept, although the confidence interval is quite wide. This suggests that, overall, they have integrated fairly well into the university; however, a larger sample size would be advisable to confirm this trend.

Linear Mixed Model with $exam\ condition$ as a grouping factor.

$$
\begin{aligned}
\theta_{\text{stress-wellbe}} &= \beta_1\,\theta_{\text{academic\_perception}} + \beta_2\,\theta_{\text{academic\_expect}} \\
&\quad + \beta_3\,\theta_{\text{workload}} + \beta_4\,\theta_{\text{social\_support}} \\
&\quad + \beta_5\,\theta_{\text{satisfaction}} + \beta_6\,\theta_{\text{motivation}} \\
&\quad + b_{\text{exam},0j} + \epsilon_{ij} \\
b_{\text{exam},0j} &\sim N(0, \sigma_b^2), \\
\varepsilon_{ij} &\sim N(0, \sigma^2).
\end{aligned} \tag{7}
$$

```
PVRE = 0.054
```

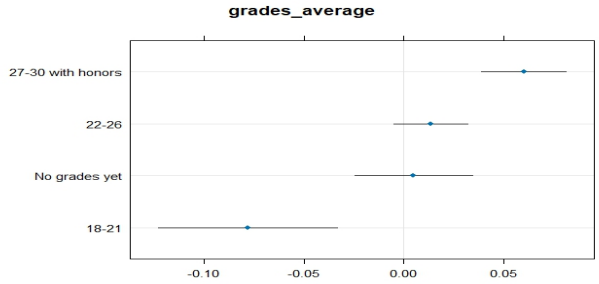

Figure 7. Random intercept plot exams average

Here, the PVRE is not as low as in other cases, standing at a solid 5.4%. Four categories, remain significant: students who are up to date with their exams (showing a positive effect), those with 1–2 backlogged exams (slightly positive effect), and those with more than three backlogged exams (showing a negative effect). Students close to withdrawal also score lower, although their confidence interval is quite wide. Overall, having backlogged exams appears to strongly influence students' stress levels.

Linear Mixed Model with $grades\ average\ (avg)$ as random intercept.

$$
\begin{aligned}
\theta_{\text{stress-wellbe}} &= \beta_1\,\theta_{\text{academic\_perception}} + \beta_2\,\theta_{\text{academic\_expect}} \\
&\quad + \beta_3\,\theta_{\text{workload}} + \beta_4\,\theta_{\text{social\_support}} \\
&\quad + \beta_5\,\theta_{\text{satisfaction}} + \beta_6\,\theta_{\text{motivation}} \\
&\quad + b_{\text{avg},0j} + \epsilon_{ij} \\
b_{\text{avg},0j} &\sim N(0, \sigma_b^2), \\
\varepsilon_{ij} &\sim N(0, \sigma^2).
\end{aligned} \tag{8}
$$

```
PVRE = 0.012
```

Once again, we do not observe a high PVRE, but a growing trend can be noticed with the grade point average, highlighting that the academic situation (in

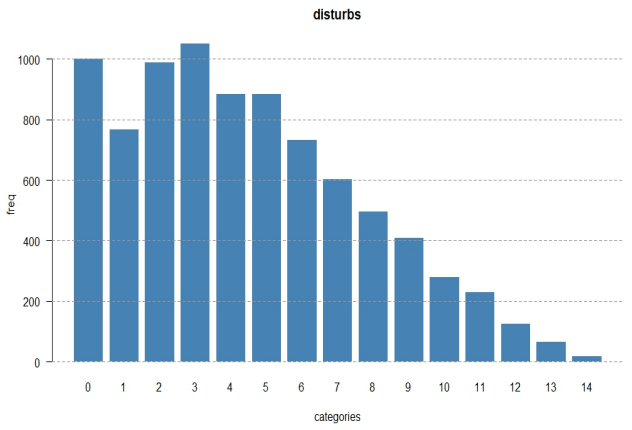

Figure 8. Frequency of the number of disturbs among students

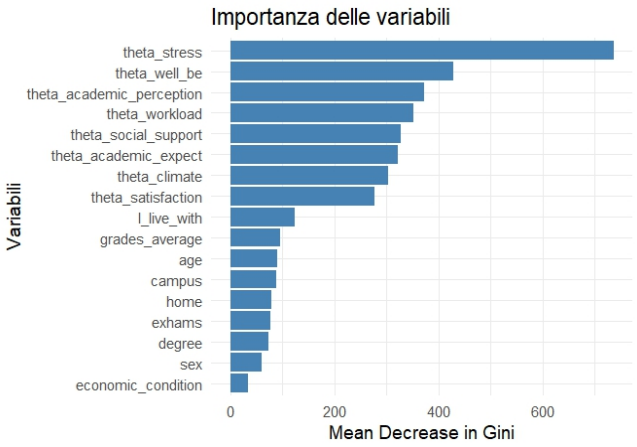

Figure 9. Random intercept plot Mod 4

this case the grade point average) turns out to be a relevant factor when it comes to student stress.

Conclusion over the risk-factors

Classification of student at-risk

A substantial part of the dataset consists derives from a survey assessing psychological symptoms that students may have experienced over the past 12 months for a period of at least two consecutive weeks. Fourteen symptoms were investigated: “Lack of interest”, “Sadness/depression”, “Low self-esteem”, “Sleep problems”, “Tension/anxiety”, “Panic attacks”, “Avoidance/escapism”, “Unexplained pain”, “Low concentration”, “Repetitive thoughts”, “Appetite problems”, “Mood swings”, “Self-harm”, “Suicidal thoughts”.

The frequency with which such symptoms were reported is very high. To provide an overview, the following figure shows the response frequency distribution:

The purpose of this part of the study is to use the latent parameters of the IRT models, previously discussed, as inputs for classification models in order to identify the students most at risk. Three risk classes were therefore created:

- LOW risk: students with fewer than 2 symptoms
- MEDIUM risk: students who reported between 3 and 5 symptoms
- HIGH risk: students with more than 5 symptoms

The objective is therefore to understand how the latent parameters, particularly in this study those from the 4PLnRM model, are able to “identify” students in significant difficulty or at least serve as an early warning signal.

The classifier used is a random forest which, in addition to its excellent predictive power, provides the variable importance score, once again useful to understand which factors most strongly affect the recurrence of these symptoms.

The model and consideration

The first model presented takes as input the same dataset used for the linear models and, as output, the vector containing the students’ risk class as previously defined.

Considering the large size of the dataset, it was split into 80% training and 20% testing; on the training set, both hyperparameter tuning of the model ($ntry = 2, 4, 6, 8$) and cross-validation were performed simultaneously to identify the best one.

mtry	Accuracy	Kappa
2	0.5721	0.3569
4	0.5829	0.3734
6	0.5789	0.3672
8	0.5789	0.3673

Table 4. Accuracy and Kappa over the train set

	HIGH	**LOW**	**MEDIUM**
HIGH	375	53	182
LOW	54	359	157
MEDIUM	152	115	204

Table 5. Confusion matrix of the test set

Statistic	HIGH	LOW	MEDIUM
Sensitivity	0.6454	0.6812	0.3757
Specificity	0.7804	0.8123	0.7590
Pos Pred Value	0.6148	0.6298	0.4331
Neg Pred Value	0.8021	0.8446	0.7127
Prevalence	0.3519	0.3192	0.3289
Detection Rate	0.2271	0.2174	0.1236
Detection Prevalence	0.3695	0.3452	0.2853
Balanced Accuracy	0.7129	0.7467	0.5674

Table 6. Statistics of the model

Conclusions on the model

The goal of this section was to verify whether the latent parameters estimated by the 4PLnRM model are able to identify and distinguish at-risk individuals from those at lower risk. It is important to note that, within the context of this study, the definition of “at-risk” is not unequivocal: a definitive classification would have required expert evaluation to objectively determine each student’s actual level of risk. Nevertheless, a good proxy is the number of symptoms experienced over the past year.

As shown, the latent parameters provide an effective tool for discriminating individuals at higher risk. Overall, the model's performance is good, though not exceptional, since—as one might expect—it struggles to clearly distinguish individuals in the "MEDIUM" class from those in the other two classes, particularly from the "HIGH" group. On the other hand, the model very accurately separates the two most distant classes, with a remarkably low error rate. This highlights how the latent parameters indeed carry meaningful information about the individuals' emotional state. This finding is further supported by the model's importance plot, which shows that these parameters rank higher than exogenous factors such as gender or age—variables that are typically considered relevant in linear models.

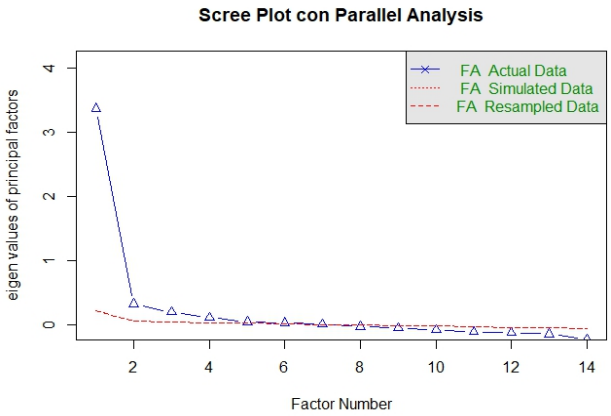

Figure 10. Scree-plot

Exploratory Factoral Analysis

introduction

As has become clear so far, in psychometrics it is often difficult to define an objective measurement scale capable of quantifying levels of discomfort/stress in a subject, and subsequently identifying the main risk factors. This section, however, focuses on symptoms experienced over the last 12 months for a prolonged period, with the aim of grouping students into risk classes that are not based merely on the number of symptoms, as was the case in the section devoted to classifying at-risk students. The idea here is to start from something more 'concrete' from a psychological perspective—such as the presence of a specific symptom—and then investigate whether there are significant differences within the population in the various previously listed risk factors (sex, age, economic condition ecc..). The method adopted for this analysis is Exploratory Factor Analysis (EFA), which, in addition to being widely used in the psychometric field (Brunner (2023),Finch (2023)), is suited to handling dichotomous data (i.e., the dataset of symptoms).

Short theory recap

In this case, the input to the model consists of the sub-dataset containing the k = 14 columns corresponding to the questionnaire items representing different domains of mental health symptoms, listed in the classification model.

The factor analysis model may be written as follows. Independently for $j = 1, \ldots, n$ subjects, let

$$z_j = \Lambda F_j + e_j$$

where z_j is a $k \times 1$ observable random vector ($z_{j1} = 1$ if subject j has the disturb 1, otherwise), Λ is a $k \times p$ matrix of constants,

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1p} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k1} & \lambda_{k2} & \cdots & \lambda_{kp} \end{bmatrix}$$

and F_j (for factor) is a $p \times 1$ latent random vector with covariance matrix Φ.

The $k \times 1$ vector of error terms e_j is independent of F_j; it has expected value zero and covariance matrix Ψ, which is almost always assumed to be diagonal.

There are no intercepts, and $\mathbb{E}(F_j) = 0$. A multivariate normal assumption for F_j and e_j is common.

The λ_{ij} values will be called *factor loadings*. They are essentially regression coefficients linking the factors to the observed variables.

The number of factors (symbolized here by p) is a fundamental property of a factor analysis model. For example, it determines the number of parameters.

It is typically very important to subject-matter experts, too. You can always get their attention by asking if something they are talking about is uni-dimensional. For example: is creativity uni-dimensional? Are political attitudes uni-dimensional (primarily just left-right)? In market research, how about attitudes toward a particular product category? In a classical factor analysis, the number of common factors is generally not known in advance; it is determined in an exploratory manner.

The first guiding principle is a piece of wisdom Kaiser (1960), who pointed out that for the typical problem involving human behavior or any other complex system, there are probably hundreds of common factors.

Including them all in the model is out of the question. The objective should be to come up with a model that includes the most important factors for the variables in the study, and captures the essence of what is going on.

Simplicity is important. Other things being more or less equal, the fewer factors the better.

Model application and results

The focus of this section is therefore to understand how the model behaves with our data, the number of factors selected, and their tangible interpretation, possibly supported by the results. Following a consultation with Dr. Filomena Di Sanzio and Dr. Elena Mazza from the Polipsi Department of the Politecnico di Milano, an analysis based on the scree-plot, $p = 4$ latent factors seemed to be the most appropriate choice to balance the amount of variability explained by the model and the interpretability of the factors.

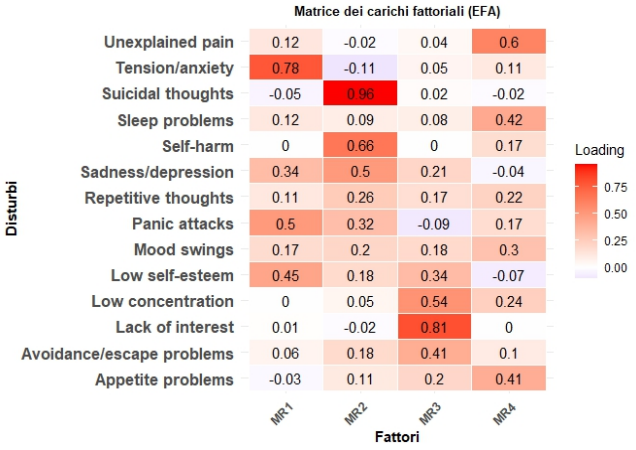

Figure 11. Factor Loadings

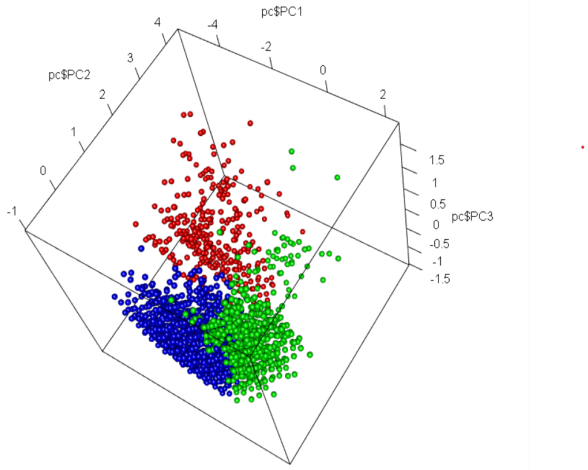

Figure 12. Cluster visualization on dimensionally reduced data (from 4 → 3)

Also reported are the factor loadings obtained from the model (11)

The factor loadings are the correlations between the observable variables and the factors. In particular, the correlation between observed variable i and factor j is λ_{ij}. The square of λ_{ij} represents the *reliability* of observed variable i as a measure of factor j.

The higher the value of a certain factor loading on a given symptom, the stronger the connection with that factor; for example, Factor 1 is strongly related to the component of tension/anxiety, but also fairly related to panic attacks and low self-esteem. Based on the results, we also provided a plausible and meaningful interpretation of these latent factors:

- MR1 = Anxiety symptomatology
- MR2 = Depressive symptomatology
- MR3 = Motivational block
- MR4 = Somatization

This is very positive because it was possible to assign a concrete meaning to the latent factors considered in our analysis, which is therefore useful to consolidate the work carried out.

Clustering analysis on EFA latent factor scores

Similarly to what happens in a Principal Component Analysis (PCA), in Exploratory Factor Analysis (EFA) it is also possible to project the data into the space of the latent factors and compute the corresponding factor scores. These scores represent the position of each subject with respect to the estimated factors, thus allowing for a reduction in data dimensionality and a more compact representation of information.

The factor scores indicate the position of each subject along the latent factors estimated by the model. In particular, high and positive values reflect a strong presence of the factor in the subject, while low or negative values indicate a placement in the opposite direction of the factor continuum. In general, the distance from zero reflects the intensity with which the factor manifests in the subject.

This procedure, in addition to enabling a dimensionality reduction (from 14 to 4), makes it possible to work with data in a continuous domain, which is often advantageous when performing cluster analysis. In fact, it is not always possible to obtain satisfactory groupings when dealing with dichotomous data.

Several clustering methods were tested, and the reported result represents a compromise between cluster performance (measured through the silhouette score index), methodological clarity, and finally, a reasonable number of clusters. It is important to note that these clusters correspond to 'risk classes' into which the subjects will be divided, so it may be reasonable to renouncing a little bit of performance to get a better interpretation.

Results

For clarity, the clustering presented was carried out on a dataset of factors score of dimension $n \times p$ ($n = number\ of\ subjects$ and $p = 4$ latent factors). The method that best satisfied the desired requirements was a hierarchical clustering with Manhattan distance and the average linkage method.

Silhouette score	0.473
Number of clusters	3
Cluster 1 size	545
Cluster 2 size	4656
Cluster 3 size	3057

Table 7. Clustering results: Silhouette score, number of clusters, and cluster sizes.

Three hypothetical risk classes appear to be a reasonable subdivision of students based on psychological symptoms, while maintaining a satisfactory clustering score that rules out a random aggregation of the data.

Furthermore, to further test the effectiveness of the clustering, we report a heatmap of symptom prevalence across the three groups in Figure 13.

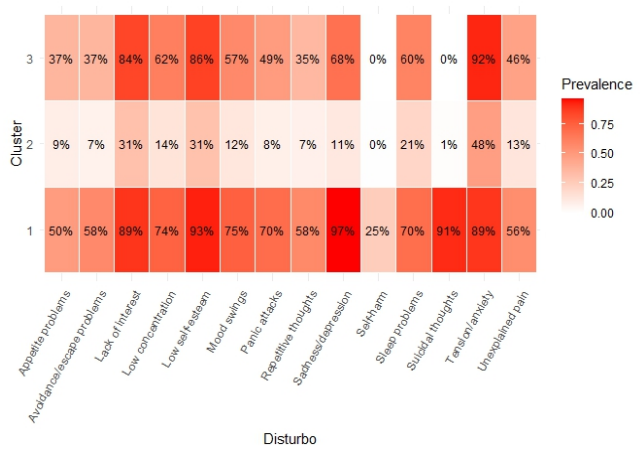

Figure 13. Symptomatic prevalences within clusters

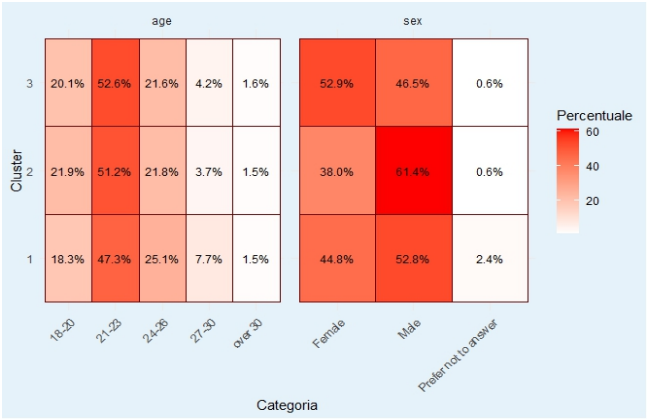

Figure 14. Differences in age and sex

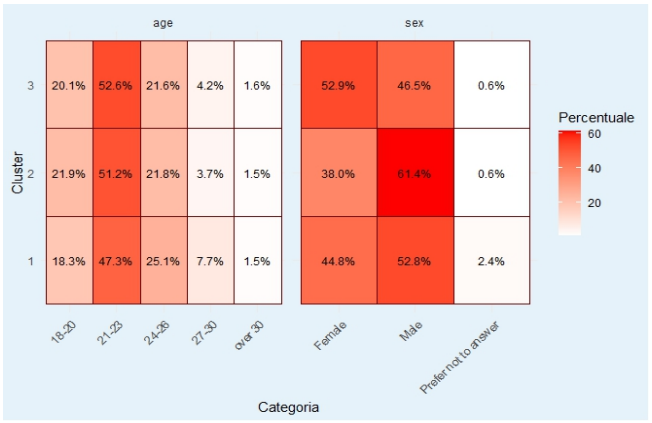

Figure 15. Differences in economic condition satisfaction and exam's average

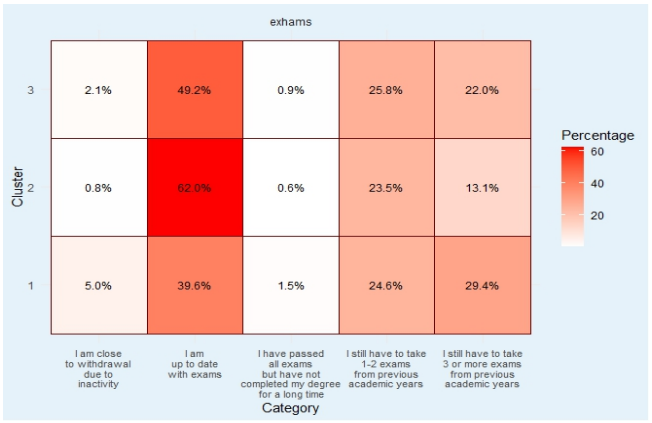

Figure 16. Differences in exam condition

From the heat map, it can be inferred that the clusters are divided into three risk categories:

- Cluster 2 (low risk) → students with few symptoms
- Cluster 3 (medium risk)→ students with several symptoms
- Cluster 1 (high risk) → students with many symptoms, including the most severe ones (self-harm and suicidal thoughts)

At a practical level, it can therefore be stated that the cluster analysis generates three student groupings with a meaningful interpretation, distinguishing them into high, medium, and low risk.

Differences in the demographic factors

Once the validity and meaningfulness of the groupings obtained from the cluster analysis on the EFA scores have been established, it is natural to ask whether there are differences in the demographic and potential risk factors considered in the linear models; this would help corroborate them as relevant factors influencing students' mental health. The results are reported in the heatmaps (figures 14 15 16).

How the graphs should be read: For the category *Exam condition* 16, for example, the rows represent the groups and the columns represent the risk factor categories. The important aspect is not the percentage itself but the difference between the clusters.

For instance, consider the percentage difference between Cluster 1 and Cluster 2 in the following categories:

- "I am up to date with exams"
- "I still have to take 3 or more exams from previous academic years"

There is clear difference among the two groups: in the Cluster 2 the percentage of students up to date with exams is high while is low on Cluster 1, while the students that still have to take 3 or more exams from previous academic years are clearly more prevalent in the first. Very significant differences can also be observed in the economic factor (the percentage of satisfaction between groups increases as we move toward the lower-risk class: 72→82→91) and in the grade-point average, but there are differences also among the age-classes (a bigger prevalence of older students in the high risk class). With regard to gender, Cluster 1 is much more balanced between males and females, whereas Cluster 2 is predominantly composed of males (this could imply that females are slightly more at risk compared to males).

In conclusion, in this part of the analysis we started from a more tangible and concrete psychological dimension, namely the symptoms experienced in the last 12 months for a period of at least two weeks, in order to group students into three risk classes. In doing so, very significant differences emerged among these groups in the various demographic factors analyzed in the previous parts of the study, revealing several consistencies with what had previously been found:

economic condition, exam backlog, and grade point average are factors that strongly impact students' mental health.

Conclusions

The primary objective of this study was to develop reliable quantifiers capable of capturing the degree of stress experienced by students and, consequently, to identify which demographic or potential stress-related factors most strongly influence it. The 4PLnRM model, thanks to its ability to capture different patterns of responses, provides these parameters (namely, the latent parameter θ).

The effectiveness of these parameters is further supported by the classification model based on Random Forests, which takes as input the estimated θ values in the various subcategories of stress, and as output an objective measure of *distress*, namely the number of psychological symptoms experienced during the last year for a prolonged period. These were grouped into three categories: high, medium, and low risk.

The classifier clearly shows, through the importance measure, the key role played by these latent parameters, in addition to the fact that its performance is highly satisfactory (especially in distinguishing high-risk from low-risk individuals). Therefore, beyond being a valid tool for building a classifier, these parameters also prove to be a reliable measure when it comes to quantifying the degree of distress in a subject undergoing such a questionnaire.

In conclusion, this evidence supports and strengthens the findings obtained from linear models: while linear approaches provide a more theoretical explanation, they alone cannot guarantee effectiveness in measuring stress levels among subjects. With greater confidence, we can therefore state that both economic and academic components play a crucial role in the mental health spectrum of students.

Finally, to gain an additional different perspective, the EFA analysis and the clusterings derived from the factor scores clearly show very significant differences within the population with respect to the same risk factors, across the three identified classes, thus further consolidating the results found in the previous sections. Moreover, rather significant differences also emerge in factors such as gender and age, although these differences are less pronounced compared to the other categories mentioned above.

These factors should thus be strongly taken into account when attempting to improve the overall academic experience at the university.

References

Bedewy D and Gabriel A (2015) Examining perceptions of academic stress and its sources among university students: The Perception of Academic Stress Scale. *Health Psychology Open* 2(2): 1–9.

Brand H and Schoonheim-Klein M (2009) Is the OSCE more stressful? Examination anxiety and its consequences in different assessment methods in dental education. *European Journal of Dental Education* 13(3): 147–153.

van der Linden WJ and Hambleton RK (1997) *Handbook of Modern Item Response Theory*. New York: Springer.

Barton MA and Lord FM (1981) An upper asymptote for the three-parameter logistic item-response model. *Psychometrika* 46: 443–450.

Rulison KL and Loken E (2009) The impact of careless responding on the estimation of item response theory models. *Applied Psychological Measurement* 33: 406–425.

Loken E and Rulison KL (2010) Estimation of item response theory models in the presence of careless responding. *Journal of Educational Measurement* 47: 205–224.

Rasch G (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.

Bech P, Allerup P, Maier W, Albus M, Lavori P and Ayuso JL (1992) The Hamilton scale and the Hopkins Symptom Checklist (SCL-90): A cross-national validity study in patients with panic disorders. *The British Journal of Psychiatry* 160: 206–211. doi:10.1192/bjp.160.2.206.

Cressie N and Holland PW (1983) Characterizing the manifest probabilities of latent trait models. *Psychometrika* 48(1): 129–141. doi:10.1007/BF02294020.

Lord FM and Novick MR (1968) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley. With contributions by A. Birnbaum.

Chalmers RP (2012) mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software* 48(6): 1–29. doi:10.18637/jss.v048.i06.

Storme M, Tavani JL and Myszkowski N (2019) Psychometric properties of nested logit models: A comparison with the four-parameter logistic model. *Applied Psychological Measurement* 43(2): 120–134. doi:10.1177/0146621618771003.

Suh Y and Bolt DM (2010) Nested logit models for multiple-choice item response data. *Psychometrika* 75(4): 454–473. doi:10.1007/s11336-010-9161-2.

Kaiser HF (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20(1): 141–151. doi:10.1177/001316446002000116.

Brunner J (2023) *Structural Equation Models: An Open Textbook (Draft)*.

Finch WH (2023) *A comparison of methods for determining the number of factors to retain with exploratory factor analysis of dichotomous data*.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

68/2025 Leimer Saglio, C. B.; Corti, M.; Pagani, S.; Antonietti, P. F.
A novel mathematical and computational framework of amyloid-beta triggered seizure dynamics in Alzheimer's disease

Leimer Saglio, C. B.; Corti, M.; Pagani, S.; Antonietti, P. F.
A novel mathematical and computational framework of amyloid-beta triggered seizure dynamics in Alzheimer's disease

67/2025 Antonietti, P.F.; Corti, M.; Gómez S.; Perugia, I.
Structure-preserving local discontinuous Galerkin discretization of conformational conversion systems

66/2025 Speroni, G.; Mondini, N.; Ferro, N.; Perotto, S.
A topology optimization framework for scaffold design in soilless cultivation

65/2025 Pottier, A.; Gelardi, F.; Larcher, A.; Capitanio, U.; Rainone, P.; Moresco, R.M.; Tenace, N.; Colecchia, M.; Grassi, S.; Ponzoni, M.; Chiti, A.; Cavinato, L.
MOSAIK: A computational framework for theranostic digital twin in renal cell carcinoma

64/2025 Celora, S.; Tonini, A.; Regazzoni, F.; Dede', L. Parati, G.; Quarteroni, A.
Cardiocirculatory Computational Models for the Study of Hypertension

63/2025 Panzeri, S.; Clemente, A.; Arnone, E.; Mateu, J.; Sangalli, L.M.
Spatio-Temporal Intensity Estimation for Inhomogeneous Poisson Point Processes on Linear Networks: A Roughness Penalty Method

62/2025 Rigamonti, V.; Torri, V.; Morris, S. K.; Ieva, F.; Giaquinto, C.; Donà, D.; Di Chiara, C.; Cantarutti, A.; CARICE study group
Real-World Effectiveness of Influenza Vaccination in Preventing Influenza and Influenza-Like Illness in Children

61/2025 Coclite, A; Montanelli Eccher, R.; Possenti, L.; Vitullo, P.; Zunino, P.
Mathematical modeling and sensitivity analysis of hypoxia-activated drugs

60/2025 Antonietti, P. F.; Caldana, M.; Gentile, L.; Verani M.
Deep Learning Accelerated Algebraic Multigrid Methods for Polytopal Discretizations of Second-Order Differential Problems