University of Naples Federico II

Department of Social Sciences

PhD Program in

Social Sciences and Statistics

Cycle XXXV

Latent Class Analysis for proficiency assessment in Higher Education: Integrating multidimensional latent traits and learning topics

Supervisors

Prof. Francesco Palumbo

Prof. Dora Gambardella

PhD student

Rosa Fabbricatore

Year 2022-2023

Acknowledgements

I want to thank all who enriched my PhD experience. First and foremost, I am deeply grateful to my supervisor Prof. Francesco Palumbo, who brought to light my interest in Statistics about ten years ago. Your knowledge and experience helped me a lot in understanding how to do research. Thanks for all your time and your continuous and valuable support. I also want to thank my second supervisor Prof. Dora Gambardella, for her interest in my research project and her availability.

Part of the thesis was developed during my PhD visiting period at Leiden University. Thus, let me express my gratitude to the Department of Methodology and Statistics, particularly Prof. Mark de Rooij and Dr. Zsuzsa Bakk, for their welcome that made my study in the Netherlands a wonderful time. Thanks for your insightful comments and suggestions that deeply improve my work. Thanks also to Dr. Roberto Di Mari, from University of Catania, who readily accepted to be involved in this research project. Thanks especially for your technical support, I learned a lot from you.

Definitely, my experience would not have been the same without the LabStat group at the Department of Political Sciences of University of Naples. Thanks to all of you for contributing to the joyful and comfortable environment that makes it possible to work with enthusiasm every day.

Thanks to my family and all my friends for their love and belief in me. Thanks, Carla, for sharing every single step of this route, doubling its value.

Contents

Introduction

The present work is devoted to the study and the methodological development of statistical approaches in the framework of Learning Analytics, with a particular focus on the issues of students' assessment, profiling and tutoring. Specifically, it proposes two statistical approaches based on latent variable models aimed to offer helpful statistical models for handling students' learning activities data or as a knowledge base in a self-recommendation learning environment. Indeed, the vast amount of data from modern technology-based learning platforms requires transforming data in knowledge as support to address the evaluation's final aim.

In this vein, statistical modeling turns out helpful for educational institutions with at least three advantages: *(i)* it offers a very detailed proficiency assessment through sets of model parameters that ensure comparison within and between students over time; *(ii)* it allows forecasting students' academic performance for the near and distant future and implementing proper actions to limit and reduce students' leaving; *(iii)* it provides insights about students' weakness that are helpful to set up tailored recommendation systems.

Conceived as a complex process, students' ability assessment accounts for: *Topics* (learning modules) of a specific knowledge domain; *Dimensions* of students' ability (specific skills); *Individual characteristics* affecting students' achievements and performance (e.g., emotional and motivational aspects). Notably, cognitive abilities independently considered from emotional, psychological, and motivational aspects cannot alone thoroughly explain learners' proficiency, academic performance and achievements (Thomas et al., 2017). Specific psycho-social aspects affect performance and achievements positively (e.g., self-efficacy, grit, positive attitude, self-regulation), whereas others have a debilitative influence (e.g., anxiety, procrastination, boredom). Therefore, students' psychological characteristics should enter as well the learning proficiency analysis, as they contribute to understand students' competencies, where the term *competencies* refers to the whole set of personal knowledge (not strictly related to the specific acquired skills), characteristics, and behaviors that lead students to success in learning.

From a theoretical point of view, students' ability can be conceived as a multidimensional latent construct measured by sets of manifest indicators. Therefore, students' proficiency evaluation depends on their responses to sets of questions that are structured within homogeneous topics. Because of the nature of not directly measurable students' ability, latent variable models stand out as an appropriate reference framework for assessing students' ability. Several statistical approaches have been proposed in this framework. Chapter 1 provides an overview of the traditional latent variable models and some of their extensions that are frequently of theoretical and empirical interest.

Statistical models play an ever more fundamental role in analyzing students' performance, especially looking at the progressive higher education unfolding to technology-driven learning (OECD, 2015; 2021). As discussed in Chapter 2, self-learning platforms have gained increasing popularity in recent years, since they constitute flexible learning environments in which technology takes over the role of teachers. Consequently, for resulting helpful to teachers and students, they must adequately analyze responses and provide adaptive feedback about topics needing review and advice about beneficial learning strategies given the students' knowledge state (Holmes et al., 2018; Fadel et al., 2019). In addition, detected information about psychological and motivational factors can drive the development of motivational feedback.

To this end, essential steps are the definition of the structure of the knowledge domain to investigate, the selection of solid criteria to assess students' ability and psychological factors, and the identification or development of appropriate indicators for their measurement. About that, Chapter 3 describes the assessment procedure specifically designed to evaluate students' proficiency in the considered application context, namely learning Statistics in non-STEM[1] degree courses. Data were collected within the "Adaptive LEArning system for Statistics" (ALEAS) ERASMUS+ Project (KA+ 2018-1-IT02-KA203-048519). In particular, data collection was carried out via the Moodle platform and consisted of three waves, each focusing on different statistical topics. The dimensions of students' ability were defined according to the Dublin descriptors, representing one of the bases for the Framework for Qualifications of the European Higher Education Area (Gudeva et al., 2012). Accordingly, a set of multiple-choice questions was developed for each considered Dublin descriptor. Finally, psychometrics scale administration allows to evaluate psychological and motivational factors hypothesised to impact students' performance.

From a statistical modeling point of view, the proposed assessment procedure requires integrating multidimensionality (more variables defining students' ability), longitudinal design with a time-varying measurement model (different topics per time point), and the covariate effects on the students' progress in learning.

The present contribution proposes two novel statistical approaches in the framework of latent variable models, which allow to manage all the above-mentioned elements effectively. Both proposals exploit non-parametric

[1]STEM states for Science, Technology, Engineering, and Mathematics

approaches that represent an ideal tool for developing accurate feedback during learning, allowing to qualify, in addition to quantify, individual differences (McMullen and Hickendorff, 2018). In particular, Chapter 4 presents a novel implementation of the multilevel latent class analysis as a classification strategy to manage complex data structures in ability assessments. As a methodological novelty, compared to previous works analyzing educational data, the proposal specifies a multidimensional latent structure at the low level of the hierarchy to account for the multidimensional nature of students' ability. Among the advantages of this approach, there is the ease of implementation and the availability of statistical software for parameter estimation.

Chapter 5 includes the second proposal of this thesis, which introduces a three-step rectangular latent Markov modeling as an extension of the traditional latent Markov models. This approach allows both for time-varying measurement models and different number of classes for the considered time points. As such, this proposal represents the first integration of the bias-adjusted three-step latent Markov modeling proposed by Di Mari et al. (2016) and the rectangular latent Markov modeling (Anderson et al., 2019). Given the originality of the proposal, no software was available for model implementation. Thus, R code for parameter estimation has been developed, and a simulation study was carried out to evaluate the performance of the bias-adjusted estimator for the third step of the proposed approach.

Note that both the proposals mentioned above also allow for considering the effect of individual characteristics on achievements and learning over time. The results from the empirical application of the proposed approaches for the analysis of the complex data structure deriving from the students' assessment proposed in Chapter 3 complement the theoretical aspects.

Chapter 1

Measuring non-observable concepts

Everyday living posits humans to face several non-observable concepts. Indeed, people are continuously called to understand, describe, and examine notions that cannot be directly measured but only inferred from a series of observable indicators. When we have to express the level of satisfaction for a product or a service, for example, our evaluation relies on a set of measurable characteristics, such as the price, the perceived quality, and the fulfillment of expectations, representing observable indicators of unobserved customer satisfaction. Also, the direct measurement of the quality of life is not possible; however, work and health conditions, income, exposure to pollution, and time spent on leisure activities can be used as indirect measures. Other examples involve emotions, abilities, personality traits, and beliefs.

Due to their unobservable nature, theoretical concepts are defined *latent* constructs, whereas the corresponding observable manifestations constitute their *manifest* proxies (Raykov and Marcoulides, 2011). The description of a latent construct in terms of manifest indicators is named *operational definition* (Crocker and Algina, 1986). Examples of manifest indicators are directly measurable properties or, as is usual in social and behavioral sciences, a set of Likert-type items. But measuring latent constructs is not an easy issue: choosing among different operational definitions of the same construct, determining an adequate sample of observable proxies, and accounting for measurement errors stand out among the major challenges (Crocker and Algina, 1986; Raykov and Marcoulides, 2011). Moreover, the complexity of some theoretical concepts requires to differentiate among their multiple facets, named *dimensions*, leading to what are known as multi-dimensional latent constructs (Rabe-Hesketh and Skrondal, 2008; Briggs and Wilson, 2003). Intellectual functioning, for example, can be considered a multidimensional construct made up of the dimensions of verbal comprehension, perceptual reasoning, working memory, and processing speed (Wechsler, 2003).

The development of psychometric theories and statistical models has primarily played an essential role in defining how observable indicators are linked to the latent construct. In the modeling approach, latent constructs are conceived as random variables with unknown values that can be inferred from observed (measured) variables throughout a statistical model (Skrondal and Rabe-Hesketh, 2007). According to the theoretical definition of latent constructs and manifest indicators, the terms “manifest” and “latent” are also used to define observed and unobserved variables in statistical modeling. Over the years, a general framework for latent variable modeling has been tuned (see, among others, Muthén, 2002; Skrondal and Rabe-Hesketh, 2004; Borsboom, 2008), aiming to gather the properties common to a wide range of latent variable models. The following sections address these similarities as well as some of the more common extensions derived from the traditional latent variable models.

1.1 Latent variable models

There are many formal definitions of a latent variable (Bollen, 2002). In the classical test theory (Spearman, 1961; Novick, 1966), for example, it represents the *true score*, namely the expected value of the observed variable for a particular individual. However, the *local independence* definition (McDonald, 1981; Lazarsfeld and Henry, 1968) is the most common definition (Bollen, 2002). Formally, let Θ denotes the latent variable and $\mathbf{Y}$ the vector of K indicators with the generic element Y_k ($k = 1, \ldots, K$). The joint density probability distribution of the latent and observed variables can be expressed as:

$$f(\Theta, \mathbf{Y}) = g(\Theta)h(\mathbf{Y}|\Theta) = g(\Theta) \prod_{k=1}^{K} h(Y_k|\Theta); \tag{1.1}$$

where $g(\Theta)$ is the probability function of the latent variable and $h(Y_k|\Theta)$ denotes the probability function of item k conditional on the latent variable. In particular, the first decomposition in Equation (1.1) accounts for the dependency of $\mathbf{Y}$ on Θ, whereas the second one specifies the local independence assumption that allows considering the K indicators as independent given Θ (Skrondal and Rabe-Hesketh, 2004). Thus, the observed covariation in the manifest variables is due to their dependence on the latent variable. It is worth noting that the local independence assumption also holds for multidimensional latent variable models, where a vector of latent variables

$\mathbf{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_D)'$ underlying the set of indicators is considered (Raykov and Marcoulides, 2011).

Starting from the general formulation in Equation (1.1), different types of latent variable models can be obtained according to the specification of the probability function of the latent variable $g(\Theta)$ and the conditional distribution of the K items $h(Y_k|\Theta)$. In addition, the response part of the model comprises also the specification of a *link function*, namely the particular regression model used to connect (hence *link*) the observed indicators with the latent variable (Bartholomew and Knott, 1999). Thus, based on the nature of manifest and latent variables, four main (traditional) types of latent variable models can be defined: factor analysis (or common factor model), latent profile analysis, item response theory models, and latent class analysis (see Table 1.1). The following subsections briefly describe the principal features of each of them.

Table 1.1: Classification of the traditional latent variable models (from Kankaraš et al., 2011).

	Latent variable	
Manifest variables	Continuous	Categorical
Continuous	Factor analysis	Latent profile analysis
Categorical	Item response theory	Latent class analysis

1.1.1 Factor analysis

The term *factor analysis* was introduced by Thurstone (1931), even though Spearman (1904) had previously referred to the common factor concept while arguing about intelligence (Skrondal and Rabe-Hesketh, 2007). However, the definition of factor analysis as a statistical method came later with Lawley (1940), Rao (1955), and Lawley and Maxwell (1971).

In factor analysis (Gorsuch, 1983), both the observed and latent variables are continuous normally distributed, and linear regression models are used as link functions. In particular, given the vector of observed variables $\mathbf{Y}$ measuring the latent variable Θ, the factor analysis model is defined as:

$$\mathbf{Y} = \boldsymbol{\lambda}\Theta + \boldsymbol{\epsilon}$$

where $\boldsymbol{\lambda}$ is a vector of factor loadings indicating the impact of Θ on the manifest indicators included in $\mathbf{Y}$, and $\boldsymbol{\epsilon}$ is a vector of measurement errors. The model assumes that the common factor Θ and the error component $\boldsymbol{\epsilon}$ have a zero mean and are uncorrelated:

$$\mathbb{E}(\Theta) = 0; \qquad \mathbb{E}(\boldsymbol{\epsilon}) = 0; \qquad \mathrm{Cov}(\Theta, \boldsymbol{\epsilon}) = \mathbf{0}.$$

It is worth noting that in literature there are two main approaches to common factor modeling: the exploratory factor analysis (Thurstone, 1935; Thomson, 1938) and the confirmatory factor analysis (Jöreskog, 1971,9). The former is based on a data-driven procedure, extracting the number of common factors from the data without specifying the loading patterns between the observed and the latent variables. Conversely, the latter defines the number of latent factors and their association with the corresponding manifest indicators according to a substantive theory or a research design before analyzing the data. For more details about the differences between these two approaches, see Hurley et al. (1997), Thompson (2004), Suhr (2006), among others.

1.1.2 Latent profile analysis

Latent profile model was firstly introduced by Green Jr (1952), even if term was later coined by Gibson (1959).

The latent profile analysis (Collins and Lanza, 2009; Peugh and Fan, 2013), as well as the latent class analysis discussed below, constitutes a person-oriented approach to latent variable analysis (Woo et al., 2018; Bergman et al., 2003), in contrast to variable-centered approaches (e.g., factor analysis). Indeed, starting from the manifest variable covariation, latent profile analysis uncovers latent groups of people, called *profiles*, instead of latent factors. The detected latent profiles represent meaningful patterns of attributes that occur across individuals (Bergman et al., 2003). The latent variable is assumed to be discrete; specifically, it has a multinomial distribution referring to the profile membership probabilities. On the other hand, manifest variables are treated as normally distributed and are linked to the latent variable through linear regression models.

For latent profile analysis, the basic local independence assumption specifies that all the K considered indicators are uncorrelated within each latent profile; thus involving profile-specific covariance matrices with all the off-diagonal elements equal to zero. In addition, parsimony issues during parameter estimation commonly require to impose also the homogeneity restriction on the variances of response variables across latent profiles (Lubke and Neale, 2006). Hence, let $\mathbf{Y}$ be the vector of observed variables and X denotes the categorical latent variable taking values $i = \{1, \ldots, I\}$, the above-mentioned assumptions can be formalized as follows:

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}; \qquad \mathbf{Y}_i \sim N[\boldsymbol{\mu}_i, \mathbf{\Sigma}].$$

Due to the local independence and homogeneity assumptions, the marginal probability density function for the generic response vector $\mathbf{Y}_s = (Y_{s1}, \ldots, Y_{sK})'$ of subject s, can be obtained following the law of total probability as:

$$f(\mathbf{Y}_s) = \sum_{i=1}^{I} P(X_s = i) h(\mathbf{Y}_s | X_s = i)$$

where I is the total number of latent profiles, $P(X_s = i)$ is the prior probability of belonging to latent profile i, and $h(\mathbf{Y}_s | X_s = i)$ the conditional probability density function of $\mathbf{Y}_s$ given the membership to the profile i.

Finally, subjects are allocated in the latent profile for which they report the highest posterior probability of membership, given their response pattern.

1.1.3 Item response theory models

Item response theory (IRT) models, often referred to as *latent trait* models, have roots in the educational field, particularly in the work of Thurstone (1925), who introduced their fundamentals to measure students' abilities. Accordingly, the latent variable is commonly called *ability* in IRT models. The introduction of the normal ogive model by Richardson (1936) and Ferguson (1942), the work of Lord (1952) on the difference between observed test score and latent trait, the definition of the Rasch model, and the related specific objectivity principle by Rasch (1960) represent other milestones in this class of models. For a more careful description of the history of IRT models, see Thissen and Steinberg (2020).

Theoretically, IRT models assume a normally distributed latent variable measured by a set of nominal or ordered categorical variables, usually called *items*. Given the multinomial or binomial distribution of the observed variables, logit or probit models are usually employed as link functions.

A peculiarity of IRT models lies in estimating the probability of an answer to a certain item as a function of both subject's ability and item characteristics. Since these models involve categorical items, it is worth distinguishing between the case of dichotomous and polytomous items. Dichotomous items are binary items assuming only two values, such as yes-no, correct-wrong, or true-false. Conversely, polytomous items have more response modalities that could be ordered, as in the case of a 5-point Likert-type item ranging from "completely disagree" to "completely agree", or unordered, as in multiple-choice questions about preferences.

Regarding the dichotomous scored items, the most general logit model formulation is the 4-parameter logistic (4-PL) IRT model (Barton and Lord, 1981), which specifies the probability that a subject s with an ability level Θ_s endorses the generic item k as:

$$P(Y_{sk} = 1 | a_k, b_k, c_k, d_k, \Theta_s) = c_k + (d_k - c_k) \frac{\exp\{a_k(\Theta_s - b_k)\}}{1 + \exp\{a_k(\Theta_s - b_k)\}},$$

where a_k is the *item discrimination*, namely the item's ability to discriminate between different levels of the latent trait, b_k represents the *item difficulty*, equal to the level of latent trait needed to have a probability $p = 0.5$ of endorsing the item, and c_k and d_k are used to identify the *guessing* (correct answer by chances) and *ceiling* (missingness for higher-ability individuals) parameters, respectively.

Reduced formulations of this model can be obtained according to the number of considered item parameters: the Rasch model (Rasch, 1960) only considers the item difficulty; the 2-PL IRT model proposed by Birnbaum (1968) accounts for both item difficulty and discrimination; the 3-PL IRT model, introduced by Barton and Lord (1981), also looks at the guessing parameter in addition to the difficulty and discrimination ones.

For what concerns polytomous IRT models, the probability that a subject s with an ability level Θ_s select the category j of item k can be generally expressed as follows:

$$\mathrm{h}[P(Y_{sk} = j|\Theta_s)] = a_k(\Theta_s - b_{jk}), \qquad j = 1, \ldots, J-1,$$

where $\mathrm{h}(\cdot)$ is the link function, J the number of modalities of item Y_k, a_i and b_{ir} the discrimination and the item-step difficulty parameter, respectively.

According to the link function specification, different IRT models can be obtained. In particular, there are three main different ways to model the comparison between ordered categories of an item: *(i)* the global logit, a cumulative approach comparing the probability that an item response is in category j or higher ($Y_{sk} \geq j$) with the probability of a response in a lower category ($Y_{sk} < j$); *(ii)* local logit, an approach comparing each category j with the previous one $j-1$; *(iii)* the continuation ratio logits, that is based on the comparison between the probability that an item response is in category j or higher ($Y_{sk} \geq j$) and the probability of a response in the previous category $j-1$. In addition, a multinomial logit can be used for nominal items, which allows comparing each category j with the reference one (say, for example, $j = 0$). IRT models relying on a global logit are called *graded response models* (Samejima, 1969, 2011), models based on a local logit are known as *partial credit models* (Masters, 1982), models exploiting a continuation ratio logit are referred to as *sequential models* (Tutz, 1990). The *nominal response model* introduced by Bock (1972) represents the most widely-used model for nominal responses. Note that several restrictions can be imposed on item parameters resulting in constrained versions of the above-mentioned IRT models. For a more exhaustive taxonomy of IRT models, see Bartolucci et al. (2015), Tutz (2020) and the references therein.

1.1.4 Latent class analysis

Latent class analysis (LCA) finds its origins in the early works of Lazarsfeld and Henry (1968), Goodman (1974), and McCutcheon (1987).

Starting from observed response patterns, LCA makes it possible to identify homogeneous groups of individuals, called *latent classes*, which define the categories of the underlying discrete latent variable. Accordingly, LCA is considered a mixture model following a person-centered approach as the latent profile analysis, with the only difference that in LCA observed variables are also categorical.

More formally, let $\mathbf{Y}_s = (Y_{s1}, \ldots, Y_{sK})'$ be the vector of responses to the K items for the subject s, and X be the categorical latent variable taking values $i = \{1, \ldots, I\}$, the basic latent class model can be defined as follows:

$$P(\mathbf{Y}_s) = \sum_{i=1}^{I} P(X_s = i) P(\mathbf{Y}_s | X_s = i),$$

where I is the total number of latent classes, $P(X_s = i)$ the prior probability of belonging to latent class i, and $P(\mathbf{Y}_s | X_s = i)$ the conditional response probability given latent class membership i.

Given the categorical nature of the manifest variables, logit or probit models are usually used as link functions in LCA with proper differences according to the binary, nominal, or ordinal scoring of the considered items. Thus, as for IRT models, binary, multinomial, or adjacent-category ordinal logistic regression can be used to model the dependency of manifest indicators from the latent variable (see, for example, Heinen, 1996; Vermunt, 2001; Magidson and Vermunt, 2004). At the end of the estimation procedure, item response probabilities are used to characterize the latent classes, whereas class membership probabilities allow assigning subjects to latent classes.

1.2 Main extensions of traditional latent variable models

Latent variable models have been very popular, especially in social and behavioral sciences. Over the years, the increasing access to a vast amount of data, and the complexity of the emerged data structures constitute new and serious challenges for data analysis, leading to the development of different extensions of traditional models to address a variety of research purposes. In the following subsections, some of these extensions are described, each contributing to the solution of one of the main issues typically arising in social and behavioral sciences investigations. A summary scheme of the discussed extensions according to the four traditional latent variable models is provided in Figure 1.1. Note that other models can be derived from the integration of two or more of the presented extensions and that the below-mentioned models merely account for some of the models of greatest theoretical and empirical interest, but it is not an exhaustive list.

1.2.1 Multidimensionality of the latent variable

When dealing with complex latent constructs defined by several more specific dimensions, unidimensional approaches to data analysis result too restrictive, causing misspecification and loss of information.

Violations of unidimensionality can be addressed following a consecutive approach (Briggs and Wilson, 2003), or adopting multidimensional approaches. The first considers the construct's dimensions as independent of each other, and thus specifies a set of corresponding unidimensional models. Conversely, multidimensional approaches allow to account for covariances between the construct's dimensions, simultaneously modeling the correlations between all the observed variables. In particular, between-item multidimensional approaches assume that each item measures only one latent variable (construct's dimension), whereas within-item multidimensional approaches allow each item to contribute to the measurement of more latent variables (Adams et al., 1997a; Wang et al., 1997).

Regarding the models based on continuous latent variables, namely factor analysis and IRT models, three main extensions have been proposed to manage multidimensionality (Reise et al., 2010). The first one, commonly called *correlated traits model*, considers latent variables as distinct correlated "primary" traits. The second one, known as *higher-order model*, adds a higher-order (general) latent variable accounting for the correlation between the considered specific dimensions. Finally, *bifactor models* (Holzinger and Swineford, 1937; Schmid and Leiman, 1957) define a latent structure where each item is related to a general factor, reflecting commonality between items, and one or more specific factors accounting for the different construct's dimensions and explaining item response variance not accounted for by the general factor. A further extension of bifactor models, which consider multiple general factors, is the *two-tier model* (Cai, 2010). Bifactor and two-tier models represent the most common way to model within-item multidimensionality. See also Bonifay (2015) for a comparative review.

A specific formulation of the above-mentioned models in the IRT framework can be found in Reckase (2009) for multidimensional IRT models, De La Torre and Douglas (2004) for higher-order IRT models, Gibbons and Hedeker (1992) and Gibbons et al. (2007) for bifactor models, Cai (2010) for two-tier IRT models, among others.

In the case of multiple discrete latent variables, a multidimensional extension of LCA was introduced by Jeon et al. (2017), who proposed the *joint latent class analysis* (JLCA) that defines different groups of individuals according to their class memberships for all the considered latent variables. As such, this model represents a mixture of discrete latent variables.

1.2.2 Repeated measures of the latent variable

In longitudinal studies, latent variables are measured over several time occasions: in such cases, interest is not only in inter-individual differences but also in intra-individual differences accounting for latent variable changing across time (Raykov, 2007).

One of the traditional approaches for handling longitudinal data considers time points nested into individuals and thus exploits *multilevel models* for their analysis (see, for example, Raudenbush and Bryk, 2002). The contributions of Muthén (1997a; 1997b), Rabe-Hesketh et al. (2004), and Skrondal and Rabe-Hesketh (2004) clearly describe the multilevel perspective in the analysis of longitudinal data. Key references for multilevel models in the IRT context are instead Adams et al. (1997b), Fox (2001), and Fox and Glas (2001). Moreover, the general multilevel latent variable modeling framework proposed by Vermunt (2003; 2008) can also be considered to manage hierarchical data structures with discrete or continuous latent variables at each level.

In addition, several approaches within the structural equation modeling framework (e.g., Bollen, 1989) have been developed to treat longitudinal data. Among them, for example, *dynamic factor analysis models* (Geweke, 1977; Molenaar, 1985) have been proposed to measure changes in the latent variables concurrently accounting for indicators' correlation within each time point and interrelation of the same indicator across time (Marsh and Grayson, 1994; Raffalovich and Bohrnstedt, 1987). Extensions to polytomous items can be ascribed to Eid (1996) and to te Marvelde et al. (2006).

The most widely used approach to analyze continuous observed variables is the *latent growth curve model* (Meredith and Tisak, 1990), whose accurate description can be found in Bollen and Curran (2006). This approach posits the existence of continuous latent trajectories capturing individuals' change over time. The characteristics of the trajectory can vary across individuals; thus, the parameters of the curve, namely the intercept and the slope, are modeled as latent variables. In particular, the intercept describes the initial level of the considered construct (e.g., traits, attitudes), whereas the slope represents the rate of change over time.

For categorical observed variables, a second-order latent growth curve model can be employed to include an IRT measurement model (Zheng, 2017; Wang and Nydick, 2020). For a deeper discussion on the latent growth curve model and its extensions and application, see Preacher et al. (2008), Kohli and Harring (2013), Cagnone et al. (2009), Gorter et al. (2020).

Noteworthy, Bollen and Curran (2004) integrated the autoregressive effect in the growth curve model leading to the *autoregressive latent trajectory* model. It represents a hybrid model that explains within-subject dependence concurrently considering individual change trajectories, as typical in growth curve models, and the persistent effect of prior values of the observed variable on the current ones (Bollen and Zimmer, 2010). In this framework, Jeon and Rabe-Hesketh (2016) proposed a variant for longitudinal binary data.

Moreover, a generalization of the autoregressive latent trajectory model, called *latent variable autoregressive latent trajectory*, looks at the autoregressive relationships between repeated latent variables rather than manifest variables (Bianconcini and Bollen, 2018). Further extensions of this model to polytomous observed variables have yet to be discussed.

In addition, also the models developed to handle multidimensional latent variables can be used to analyze latent traits change over time. Indeed, longitudinal measurements of a latent variable can be modeled by considering a series of primary latent variables, one for each time point, whose correlation reflects the stability over time (Cai, 2010). More complex multidimensional models, such as bifactor and higher-order models, can also be used to analyze longitudinal data, given that the different measurement occasions constitute the construct's facets (Koch et al., 2018). The longitudinal bifactor IRT model (Hill, 2006) is an example referring to the IRT framework. Moreover, several other IRT models have been proposed to examine change over time. One of the earliest models was introduced by Andersen (1985), who consider the repeated administration of the same items over time, with constant item difficulties and time-varying individual ability. In particular, in Andersen's model, the change over time is measured as difference of time-specific abilities. Andrade and Tavares (2005) extended this model, defining a multivariate normal distribution for the latent variable so to obtain a covariance matrix to study the change over time. Another notable extension is the *multidimensional Rasch model for learning and change* developed by Embretson (1991), which assumes a different ability for each time point allowing also for time-varying observed variables. A version of this model for polytomous items can be found in Fischer (2001). For details on Embretson's model and on longitudinal IRT models see Von Davier et al. (2011), Wang et al. (2016), Embretson (1991).

Concerning discrete latent variables, longitudinal extensions account for transitions across latent classes, or profiles, over time. In particular, the *latent class transition model* extends the latent class model to longitudinal design, whereas the *latent profile transition model* represents the longitudinal version of the latent profile model (Hickendorff et al., 2018). Regardless, both extensions can be generally defined as *latent transition models* (Collins and Wugalter, 1992; Collins and Lanza, 2009; Chung et al., 2005), and the corresponding latent classes or profiles are usually called *latent states* to underline their temporary nature (e.g., Collins and Lanza, 2009). Note that latent transition models are also referred to as latent Markov models or hidden Markov models (Kaplan, 2008; Vermunt et al., 2008), since the latent process typically follows a first-order Markov chain, where change only depends on the previous class. Accordingly, three types of parameters characterize latent transition models: *(i)* initial state probabilities, namely state proportion at the first time point; *(ii)* transition probabilities, describing the transition from one state to another at each subsequent time point; *(iii)* class-conditional parameters accounting for the relation between latent states and observed indicators and thus characterizing the latent states (Hickendorff et al., 2018). Further details can be found in Song et al. (2017), Raykov (2007), Vermunt (2010b), Bartolucci et al. (2014b). For a comprehensive overview of the latent Markov models see Cappé et al. (2009), Zucchini and MacDonald (2009), Bartolucci et al. (2012).

1.2.3 Individual covariates affecting the latent variable

Individual covariates in latent variable models can affect the distribution of the latent variable and/or the response variables. The first case typically looks at the effect of individual characteristics (e.g., socio-demographic information) on the considered latent construct's levels, whereas the latter represents a useful tool for studying differential item functioning (Moustaki, 2003; Vermunt and Magidson, 2021; Bakk and Kuha, 2021). In this section, the covariate effect on the latent variable is mainly discussed.

The estimation of covariate effects can follow a one-step (global) approach or a multi-step (consecutive) approach where the measurement model is firstly fitted, and then the obtained factor scores or latent classes/profiles are used as observed dependent variables in a subsequent regression analysis (Moustaki and Knott, 2000; Goldstein, 1980; Zwinderman, 1991; Hoijtink, 1995). Because the consecutive approach leads to biased estimates (Croon and Bolck, 1997; Bakk and Kuha, 2021), some correction methods have been proposed especially in the latent class framework, where the hard classification of subjects disregards the classification uncertainty (see, for example, Bolck et al., 2004).

On the other hand, the global approach allows simultaneously estimating model parameters and covariate effect, resulting in what is defined *latent regression model* where the dependent variable is latent rather than directly observed. Related references can be found in Sammel and Ryan (1996) for factor analysis, in De Boeck and Wilson (2004), Rijmen et al. (2003), Zwinderman (1991) for IRT models, and in Dayton and Macready (1988), Hagenaars (1988), Muthén (2004) for latent class/profile models. Note that in the continuous case, the effect of covariates on the latent variable is modeled through a linear regression model, whereas a logit-type regression model is specified to express the covariate effects on latent class membership probabilities in latent class/profile models (Bandeen-Roche et al., 1997; Tay et al., 2011; Hagenaars and McCutcheon, 2002; Maij-de Meij et al., 2008).

In the recent contribution by Bakk and Kuha (2021), the authors outline the recommended modeling approaches for different circumstances involving latent class models with external variables.

Among the global approaches, the *multiple indicators and multiple causes* (MIMIC) model has been proposed to account for the direct and indirect (via the latent variable) effect of covariates on normally-distributed indicators (Jöreskog and Goldberger, 1975). This model has been extended to other types (i.e., binary and ordinal) of observed variables (Muthén, 1989), to the IRT framework (Bertaccini et al., 2013), and to latent class analysis (Muthén et al., 1998; Yang, 2005).

Other interesting references about statistical modeling of covariate effect on latent variables are in Vermunt (2010a), Moustaki et al. (2004), and Wang and Wang (2019).

1.2.4 Distal outcomes of the latent variable

External variables can act not only as predictors of latent variables but also as distal outcomes. As argued in the previous subsection about the effect of covariates, both a one-step and a multi-step approach can also be followed for models including distal outcomes. Moreover, the observed or latent nature of distal outcomes also determines the choice of the proper statistical model.

When the aim is exploring hypothesized connections among a set of latent variables, the structural equation modeling framework (SEM; Bollen, 1989; Bentler, 1995) stands out as a reference among the one-step approaches. Firstly introduced for continuous observed and latent variables, SEM models have been later extended also to categorical indicators (Edwards et al., 2012; Bovaird and Koziol, 2012) and categorical latent variables (Jedidi et al., 1997; Dolan and van der Maas, 1998; Vermunt and Magidson, 1922; Bauer and Curran, 2004; Arminger et al., 1999). Another example of a one-step approach for modeling the effect of a latent variable on distal outcomes can be found in Smid et al. (2020).

A two-step approach to SEM has been introduced by Anderson and

Gerbing (1988), which firstly accounts for the goodness of the measurement model through a factor analysis, and then estimates the structural connections among the considered constructs performing a path analysis. On the other hand, multi-step approaches for categorical latent variables are discussed in Bakk et al. (2013), Zhu et al. (2017), Bakk and Kuha (2021), Dziak et al. (2016), where the estimated class memberships by means of a latent class/profile analysis are used for further statistical elaborations.

Distal outcomes have been widely studied in the latent class modeling framework. In addition to the categorical SEM, that allows estimating the association between independent and dependent categorical latent variables, global approaches for categorical latent variables also include the *latent class analysis with distal outcomes* (Lanza et al., 2013; Bolck et al., 2004), which accounts for observed dependent variables. Moreover, latent class models including observed or latent continuous variables as distal outcomes have also been developed as multi-step approaches (Bakk and Vermunt, 2016; Shin et al., 2019).

Methodological studies on distal outcomes in latent class models can be found in Bakk and Vermunt (2016), Bray et al. (2015), and Di Mari et al. (2020). Some recommendations about the best modeling approaches for different situations involving distal outcomes in latent class analysis can be found in Nylund-Gibson et al. (2019) and Bakk and Kuha (2021). The paper of Feingold et al. (2014) provides an overview of the described approaches.

1.2.5 Clustered data

Clustered data are characterized by multiple levels or units of analysis. One of the widely common example regards students nested within classrooms.

Moreover, hierarchical data structures can result from complex sample designs, such as multistage sampling that uses clustered observations at each stage.

There are mainly three alternatives to deal with clustered data, each driven by a different research interest. When the focus is on the low-level unit of analysis (e.g., students), several statistical methods can be used to adjust parameter estimates according to the group-level structure, namely accounting for the dependency between observations belonging to the same group (Stapleton, 2006; Huang, 2016). The second strategy consists of aggregating individual-level data into the group level, and thus carrying out the analysis on the higher-level units (Stevens, 2012). Note that the consequent drawback resides in discarding low-level variability, so this strategy is appropriate only when the interest is in the group-level differences (e.g., between schools). Finally, *multilevel models* allow to concurrently account for both level units, decomposing the total variance in within-group and between-group variation. Multilevel models are particularly useful for addressing research questions regarding the influence of the group-level units on the low-level units' outcomes (e.g., school effect on students' performance). A discussion on the multilevel and single-level models for measured and latent variables with clustered data can be found in Stapleton et al. (2016).

Regarding multilevel factor analysis (Muthén, 1991; Reise et al., 2005; Kamata et al., 2008), the contribution of Kim et al. (2016) offers a review of the common practices and guidelines for users. As underlined by the authors, multilevel factor analysis also serves to investigate the psychometric properties of scales (see, for example, Huang and Cornell, 2016).

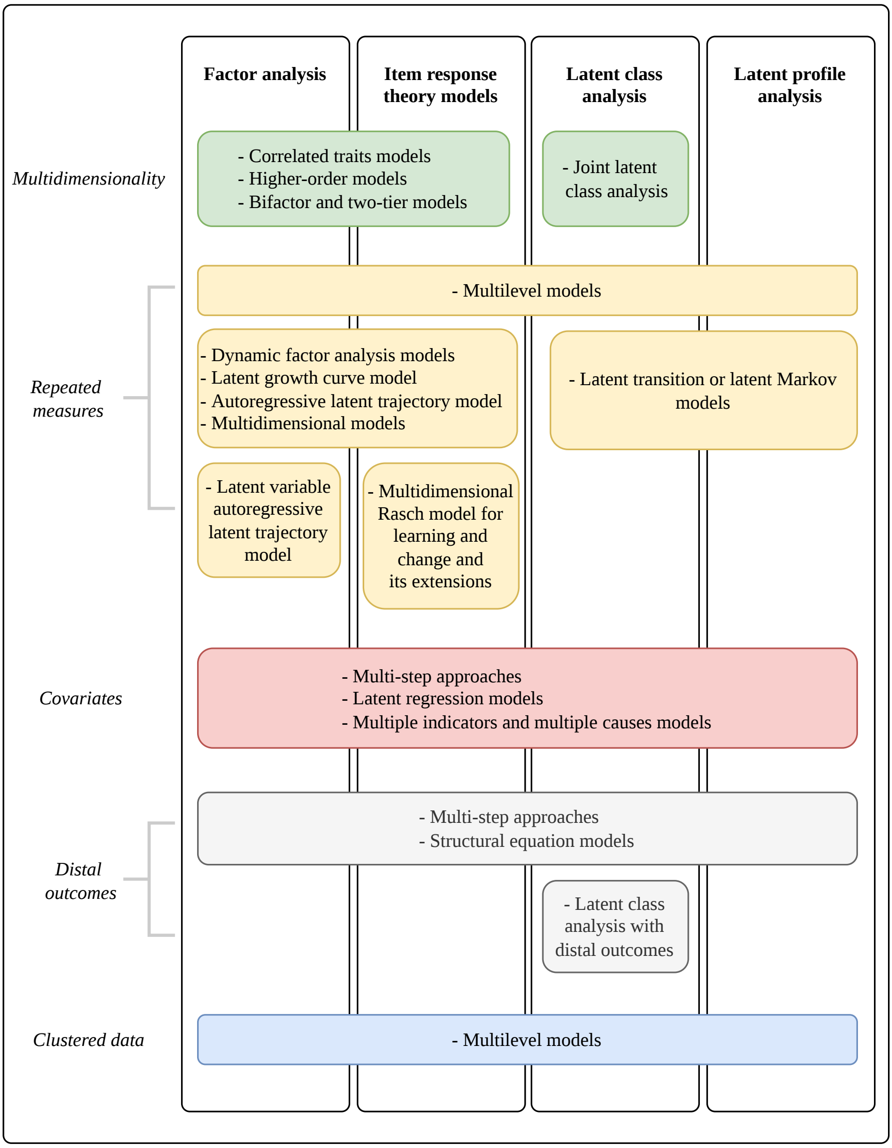
Concerning the multilevel formulation of IRT models, Fox (2001), Fox and Glas (2001), Kamata and Vaughn (2011), and Pastor (2003) represent the primary references. Multilevel IRT models comprise a "standard" IRT component for the conditional probability of the observed item responses given the latent trait along with regression component for the latent trait parameter to account for the hierarchical structure of data.

In addition, a number of multilevel latent class models have been described in the literature to account for dependency between individuals in nested data structures where discrete latent variables are considered. Vermunt (2003) proposed parametric and non-parametric models, all allowing for different parameters across higher-level units. For example, the probability of belonging to the latent classes defined at the low level can vary across groups specified at the high level. A comprehensive description of these models is given by Finch and French (2014), Henry and Muthén (2010).

Multilevel latent profile analysis (Vermunt, 2003; Asparouhov and Muthén, 2007) represents a particular version of the multilevel latent class analysis accounting for continuous indicators, thus considers, for example, within-group variations in the proportional distributions of individual-level profiles (see Mäkikangas et al., 2018 for more details about different types of multilevel latent profile models).

Other contributions also consider multilevel cross-classified structures, where lower-level units simultaneously belong to two (or more) higher-level clusters. For a description of the cross-classified multilevel models, refer to Beretvas (2011), Raudenbush and Bryk (2002), Fielding and Goldstein (2006), Rasbash and Goldstein (1994).

Figure 1.1: Summary scheme of the main extensions of traditional latent variable models.



Chapter 2

New educational frontiers in the digital era

Digital transformation is an ongoing process that continually changes the ways people do things, including learning activities. In the educational context, there is a growing interest in using technology to ensure learning achievements in a globalized society (Khatun, 2019). The OECD (2015; 2021) has also increasingly encouraged the development of digitalized learning environments that enable lifelong learning beyond the boundaries of space and time. In higher education (HE), the creation and diffusion of massive open online courses (MOOCs), as well as the implementation of e-learning and blended learning, represent relevant attempts to meet this increasing need (Buhl and Andreasen, 2018).

In this vein, the extensive implementation of technological devices during the Covid-19 emergency also accelerated the progressive HE unfolding to digital innovation, leading the way toward a new era for HE based on more sustainable approaches which encompass technology-driven learning (Ruiz-

Mallén and Heras, 2020; Bacci et al., 2022). Among them, self-learning platforms represent a very flexible learning environment that fosters and supports students' self-learning abilities. Indeed, they enable students to decide how, where, and when to learn and review their learning results.

The continuous innovations in advanced learning technologies have provided new opportunities for students, adapting learning activities to their ability level, characteristics, mood, and emotion (Harley et al., 2017). Different levels of automation could be adopted in learning technology, which integrates human and technological involvement to a different extent (Molenaar, 2021). When full automation is chosen, as often applied in self-learning platforms, students' learning paths are totally technology-driven. Thus, artificial intelligence and statistical-based technologies take over the teachers' role, resulting in what is known as an intelligent tutoring system (ITS). Over the years, ITS has increasingly improved education efficacy ensuring large-scale personalized learning where students receive problems, learning materials, and feedback tailored on their current ability level, and psychological state (Holmes et al., 2018, Holmes et al., 2019).

Several studies have shown the relevance of one-to-one compared to one-to-many tutoring, shedding light on the need for technology-based platforms to assist traditional learning methodologies (see, for example, Elsom-Cook, 1993). The adoption of ITSs can lead to improved learning outcomes compared to other teaching methods (Akyuz, 2020; Mousavinasab et al., 2021; Paviotti et al., 2012) since the one-size-fits-all curriculum provided in the standard classroom approaches may not meet students' specific learning needs. Therefore, in recent years, ITSs that collect and analyze responses during user interaction for an automated assessment and profiling were developed as a new standard to improve the learning outcome.

2.1 Self-learning platform rationale

The diffusion of self-learning platforms meets the increasing need for large-scale open learning environments and for improving learners' self-regulation skills, constituting key features of lifelong learning.

Self-regulated learning can be described, as the name suggests, as an active process whereby learners autonomously direct, monitor, and regulate their learning. According to Zimmerman (2000), self-regulation refers to "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (p.14). Thus, it implies several self-regulatory processes such as planning, cognitive activation, metacognitive monitoring, and reflection on the outcomes (Azevedo, 2009; Pintrich et al., 2000).

Self-regulation can be promoted in diverse ways in self-learning platforms. Personalized feedback, for example, can activate a reflection about own strengths and weaknesses as well as suggest more appropriate learning strategies. Moreover, self-regulation also depends on learners' engagement and motivation in learning activities (Azevedo, 2005; Veenman, 2007), which can be likewise increased through the customization of the learning process. Accordingly, to improve learners' ability and support their self-regulation skills, ITSs should address three essential elements in a self-learning environment (Molenaar, 2021): *(i) Detect* students' abilities, characteristics, and behavior during learning (Baker and Inventado, 2016; Azevedo and Gašević, 2019); *(ii) Diagnose* learners' states throughout the learning process, namely tracking their ability progress (Desmarais and Baker, 2012; Bosch et al., 2015) and changes in their emotions and motivations, identifying gaps and anticipating future development; *(iii) Act* to provide a personalized learning path following the most appropriate strategy given the learner state (Shute and Zapata-Rivera, 2012). Regarding the latter point, adaptive learning systems can enact three main actions that operate at step, task, and curriculum level, respectively (VanLehn, 2006). More in detail, providing feedback at the step level consists of explaining how to solve problems or subproblems. The second type of action regards the selection of the most appropriate next task according to the previous student's performance to best fit the student's current ability and need. Finally, with the curriculum action, the ITSs select the knowledge domains that the student is ready to learn based on the current knowledge subsets the student has already mastered.

Thus, in sum, ITSs implemented in self-learning platforms allow to customize: *(i)* the *learning tasks*, matching the difficulty level of the task and the knowledge state of the student; *(ii)* the *learning time*, giving each student the necessary time to learn a topic; *(iii)* the *learning feedback*, providing students with advice about problem-solving steps and adopted learning strategies; *(iv)* the *learning pathway*, assigning the topics in a personalized order and also informing about the topics that need to be repeated.

In this framework, it is evident that an accurate assessment of the current state of the student is a considerable part of a personalized learning activity over time. As discussed in the following subsections, these issues involve pedagogical theory since they require an understanding of how learning takes place, how to support it, and what factors have a relevant influence. In particular, herein, a social-constructivist view is presented since it is one of the most widely adopted pedagogical paradigms in self-learning platforms (Ouyang and Jiao, 2021; Secore, 2017).

2.1.1 Assessment of learners' ability

Evaluating learners' competencies is a crucial concern in education for its empowering role in learning. Students' ability assessment, conceived as a complex process, serves multiple functions: *(i)* as a tool to measure students' achievement (*assessment of learning*), *(ii)* as feedback for teachers to fine-tune their future teaching strategies to support student learning (*assessment for learning*); *(iii)* as helpful information enabling students to identify their strengths and weaknesses and monitor their progress (*assessment as learning*) (Toomaneejinda, 2017).

Thus, assessment should not be considered only as a final step of the teaching/learning moment to measure students' achievement. Instead, assessment should assume a dialectic relation with teaching and learning, consisting of an ongoing interaction that enhances teaching and learning strategies. In this vein, "testing not functioning as a gatekeeper but as a door-opener" (Toomaneejinda, 2017). Indeed, the assessment moment contributes to defining what Vygotskij called the "zone of proximal development" (Vygotskij, 1978), orienting teaching/learning activities accordingly. This integrated design significantly improves the effectiveness of the learning process, establishing if a student is ready to progress in her/his curriculum.

These principles can also constitute the foundation for designing self-learning platforms (Fadeev et al., 2019), where, as stated before, ITS technology could take over the role of teachers. Indeed, ITSs can be able to understand students' need and modulate the appropriate tasks and topics based on their current state (VanLehn, 2011). In this vein, ITS aims to provide learners with tasks that suitably challenge their ability levels and thus are neither too hard nor too easy. Indeed, exceedingly challenging tasks could make students feel anxious or frustrated, whereas too easy ones could cause boredom (Slavin, 2019). On the other hand, when tasks fall in the zone of proximal development, students are expected to be engaged in their self-learning process while working on the platform.

Moreover, as Vygotskij (1978) pointed out, any type of learning happens through mediation. In particular, two main kinds of mediators act in the process of learning: symbolic and human. The first occurs when outer signs are used to learn and remember, whereas the second refers to "a more knowledgeable other" who plays a scaffolding role when moving from the current state of knowledge to the zone of proximal development.

In self-learning platforms that embed ITS, both mediations are done by means of technologies and are based on the continuous interaction of the learner with the platform. On the one hand, digital learning environments provide tests and learning materials through different digital artefacts (e.g., 3D animations, videos, pieces of music, videogames, vignettes), exploiting a greater number of outer signs with respect to the traditional setting (Fadeev et al., 2019). On the other hand, ITS also adaptively supports students during the learning process, tailoring tasks and materials according to the students' knowledge and responding to their behavior in the digital learning environment through dynamic feedback; thus, it acts as a learning facilitator or expert scaffolding (Gadanidis, 2017; Wang et al., 2017).

Moreover, the design of a modern and performing ITS should integrate cross-sectional and longitudinal data to adapt the system to the specific user and personalize the learning experience providing accurate feedback. In particular, cross-sectional data could be helpful to provide an assessment of the student in comparison with the reference population (*normative assessment*), namely identifying students' strengths and weaknesses (as well as psychological states) in comparison with their peers. Longitudinal data allows understanding students' progress over time (*ipsative assessment*) and thus their learning outcomes; it compares a student's current state with the sequence of the previous ones. Reasonably, both these types of assessment are beneficial when developing adaptive feedback (Seery et al., 2017).

2.1.2 How do psychological factors matter?

It is well established in the literature that cognitive abilities cannot explain learners' academic performance and achievements thoroughly (Thomas et al., 2017). Indeed, many emotional and motivational aspects contribute to students' learning process and performance (see, for example, de Barba et al., 2016). Some of them affect performance and achievements positively (e.g., self-efficacy, grit, positive attitude, engagement), whereas others have a debilitative influence (e.g., anxiety, procrastination, boredom). It is worth considering the effect of psychological variables in understanding learners' proficiency, at least for two main reasons. Firstly, it allows avoiding biased evaluations of students' proficiency: for example, a bad performance may be due to a lack of motivation instead of low ability. Secondly, it allows teachers to personalize their support for students, accounting for their psychological characteristics. Indeed, a learning environment should provide students with a "zone of proximal development" not only from a cognitive perspective but also from an affective one (Murray and Arroyo, 2002).

In the context of self-learning platforms, this means that ITSs should also address students' psychological characteristics, affective states, and emotions to supply more advanced forms of personalization (see, for example, Arroyo et al., 2014; Harley et al.,2015). Accordingly, if the attention toward non-cognitive factors is minor, it is more likely that students become disengaged in learning activities and gain little from them (Kang et al., 2021). In this vein, ITS should identify short-term students' experiences, such as engagement, affect, and emotion, and more stable characteristics like motivation and interest (Walkington and Bernacki, 2019). For example, engagement and self-regulation play a primary role in determining students' effort to continue learning over time. In particular, self-regulated students better manage their learning activities (Kocdar et al., 2018), using cognitive strategies to study a specific topic and metacognitive processes to monitor their learning.

It is worth noting that short-term students' experience, as engagement, emerges mainly because of their interaction with the learning environment (Christenson et al., 2012). Hence, the development of ITSs able to promote engagement, and in general psychological states positively affecting learning, could improve self-learning platform effectiveness. In this regard, it could be helpful to assess these psychological and emotional factors through self-report measures (psychometric scales), or derive them from the students' interaction and behavior within the self-learning platform/system. About the latter, for example, students' engagement could be evaluated by considering the regularity of students' activities, rate of return, and attempts. In contrast, procrastination could be evaluated considering the time delay from the day of the first activity (Carannante et al., 2019).

Once students' characteristics and psychological state are recognized, recommendations and tailored feedback could be a good strategy to enhance students' effort in learning and, in general, to strengthen their self-regulation skills. This approach has been defined as "reactive" since the system dynamically responds to students' cognitive and psychological state during the learning progress (D'Mello and Graesser, 2014). On the other hand, the "proactive" approach is based on tools like gamification, cartoons, and other media, allowing to design platforms that intrinsically promote students' engagement and positive feelings for learning (D'Mello, 2021). Of course, the two approaches can be integrated.

According to the reactive approach, formative and motivational feedback could be supplied to improve students' cognitive performance and increase their engagement and motivation. As discussed in Hatziapostolou and Paraskakis (2010), formative feedback needs to be clearly defined and directly related to the assessment criteria. Thus, it should summarize learning outcomes and focus on the gap between the performance and the expected achievement, enabling students to identify their strengths and weaknesses and the topics they need to repeat. Therefore, this formative feedback has a developmental focus (Lizzio and Wilson, 2008), defining goals and suggesting learning strategies. On the other hand, motivational feedback aims to encourage students, recognize effort, acknowledge achievements, and give hope about their future outcomes. Thus, motivational feedback plays an important role, especially when students feel unconfident, discouraged, and anxious. Both types of feedback provide cognitive and metacognitive scaffolds to learners helping them to get back on the learning path when ineffective or inefficient strategies are used or when disengagement occurs, or anxiety and frustration take over.

2.2 The role of psychometric theories and statistical models

The widespread introduction of self-learning platforms led to the availability of large sets of data. But machine-readable data are not helpful per se, they have to be appropriately handled to produce insights for interventions.

As a consequence, a new significant area of technology-enhanced learning research emerged, called *learning analytics* (Ferguson, 2012). The Society for Learning Analytics Research (SoLAR) defined learning analytics as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Conole et al., 2011).

As such, learning analytics assumed a pivotal role in extracting information from large datasets in order to develop recommendation engines in education. According to the Learning Analytics Cycle proposed by Clow (2012), the learning analytics process starts with learners, who generate data that are processed into metrics; then, interventions are developed based on the obtained metrics, which affect learners' outcomes closing the cycle.

Practically, learning analytics can provide estimates of students' current knowledge or mental state, likely outcomes, activity on the platform, progress over time, dropout risk, and task information such as topic's difficulty. Hence, interventions can include extra help during exercise, adaptive feedback, task personalization, and motivational support (Chatti et al., 2012).

Regarding the metrics, they can be obtained by exploiting several statistical methods and other similar techniques, such as machine learning, which are able to detect meaningful patterns hidden in the data (Muslim et al., 2020). In a literature review on learning analytics, Chatti et al. (2012) pointed out classification and prediction algorithms, clustering methods, information visualization, data mining, and social network analysis as the most used techniques.

As previously stressed in Chapter 1, latent variable models represent a relevant reference framework in this context since students' ability can be conceived as a latent construct measured by a set of manifest indicators. In particular, learners' ability evaluation in self-learning platforms is typically based on their proficiency in addressing diverse types of tasks, such as single-choice tests, with a different difficulty level and referring to several skills and topics. Hence, multidimensional latent variable models are particularly useful for addressing the challenging task of assessing students' skills in a multidimensional way. Moreover, other latent variable models' extensions allow accounting for the effect of individual covariates (e.g., cognitive and psychological factors) on students' performance, which contributes to the development of motivational reinforcement. Finally, latent variable models for longitudinal data analysis permit to consider the dynamic nature of students' proficiency, behavioral and psychological states during learning. In this regard, an interesting recent work of Kang et al. (2021) describes a latent variable modeling approach to tracking flow during artificial tutoring.

It is worth recalling that statistical approaches proposed in this framework mainly belong to parametric or non-parametric approaches. Parametric approaches are usually employed when the main aim is to rank individuals, placing them on a continuous low-to-high scale. Non-parametric approaches define classes of individuals which are internally homogeneous according to the considered latent trait; as such, they allow for catching also qualitative differences between individuals, distinguishing specific response profiles. This non-parametric approach represents a helpful tool for developing tailored recommendations and remediations according to students' profiles.

What here matters to point out is the role that latent variable models have in students' assessment and feedback tuning, in addition to the widely used machine learning algorithms. As underlined in Bartolucci et al. (2018), the latent variable model framework integrates the basic principles of Statistics with the estimation procedures developed in machine learning. The adoption of statistical models involves several advantages, such as the formulation of appropriate assumptions to account for data complexity and the possibility of taking parameter and model uncertainty into account. Moreover, statistical modeling allows researchers to fulfil issues as the data generation process, and causal relations between variables, that constitute essential elements in acquiring knowledge on phenomena and driving decision-making processes.

2.3 Beyond students: Insights for teachers, educational institution, and policymakers

Self-learning platforms are mainly oriented toward learner-centered objectives. Nevertheless, these platforms are of interest to at least three other groups of stakeholders, namely teachers, educational institutions, and policymakers (Muslim et al., 2020). Indeed, metrics derived from self-learning platforms can support decision-making processes at different levels. For example, self-learning platforms can be used as a complement to traditional courses. In such a case, teachers can use learning analytics results to reflect on their teaching practices and promote actions allowed to address students' needs, improving course effectiveness (Chatti et al., 2012).

On the other hand, benefits for educational institutions include the identification of students at risk to provide interventions for reducing students' dropouts (Jayaprakash et al., 2014) and a better understanding of strategies to improve students' achievements and performance. Moreover, learning analytics can support educational institutions in adjusting course structure and making financial decisions (Chatti et al., 2012; Graf et al., 2011).

Finally, learning analytics represents a helpful source of information for policymakers, contributing to revealing the impact of technology-driven learning activities in education. Accordingly, insights for future policy directions can emerge to improve educational outcomes and enhance students' engagement. Starting from an evaluation of current practices, obtained results, and potential barriers, educational policies can embrace, for example, more investments in technological platforms for learning as well as programs to facilitate access to knowledge and reduce the digital divide. Also, an increasing integration of self-learning platforms in traditional teaching settings can help foster technological literacy, concurrently favoring the development of students' self-regulation skills. Policymakers should look at how self-learning platforms can facilitate the achievement of educational goals, using learning analytics to investigate potentials according to different contexts and purposes. Sometimes, also at the lower level, specific educational aims can be achieved by teachers and educational institutions through technology-based learning platforms by self-defining a set of rules allowing to customize contents, feedback for students, and the kind of indicators and visualization of the results (Pardo et al., 2018; Muslim et al., 2016).

2.4 Self-learning platforms for frightening subjects: learning Statistics in non-STEM degrees

Self-learning platforms able to personalize learning activities according to individual characteristics have proven to be a valuable tool to help students facing with subjects perceived as frightening (Hsu et al., 2021; Lambert and Alony, 2018). Among them, Statistics for students enrolled in non-STEM degrees stands out. Accordingly, some specific examples of technology-based learning environments developed in this context follows.

Many students who attend non-STEM programs, such as Psychology, Sociology, Medicine, and Political Science, feel unconfident and discouraged when learning Statistics, and exhibit higher statistical anxiety levels that negatively affect their achievement in Statistics (Walker and Brakke, 2017; Hanna and Dempster, 2009; Tremblay et al., 2000).

The relevance of the effect of these negative emotions on academic performance and well-being led researchers to explore innovative teaching techniques to increase motivation and improve students' statistical learning. Among them, there is the use of real-life data, active learning activities, humorous cartoons (Lesser and Pearl, 2008), and gamification (Legaki et al., 2020). In this regard, it is evident that technological platforms for self-guided learning of Statistics play a key role, providing students with personalized learning paths. Indeed, it is important to consider that each student approaches Statistics differently, varies in psychological characteristics, and feels several kinds of emotions. Accordingly, statistical education through self-learning platforms has increased over the last years. In this vein, Albert et al. (2020) provided an overview of several MOOCs developed to learn Statistics, describing different approaches and outlining challenges and opportunities. Another example of an online learning environment is the Shinyapp platform developed by Potter et al. (2016). In contrast, López Lamezón et al. (2018) demonstrated the advantages deriving from using a virtual environment to teach Statistics in Medicine degree courses. Finally, more sophisticated technologies consist of apps using animated agents to deliver learning material, as the Multimedia Probability and Statistics System (MMPASS) (Krishnasamy et al., 2020).

Among the applications specifically addressed to students enrolled in non-STEM courses, there are the Assessment and Learning in Knowledge Spaces (ALEKS; Doignon and Falmagne, 1999), an adaptive learning system based on Knowledge Space Theory (Doignon and Falmagne, 2016), and

Stat-Knowlab (de Chiusole et al., 2020), a web-app that provides students with tailored learning exploiting a competence-based extension of knowledge space theory (Heller et al., 2015).

Finally, a brief description of the mobile app named Adaptive LEArning system for Statistics (ALEAS; Fabbricatore et al., 2021; Pacella et al., 2022) follows. The system was developed in the contest of the ERASMUS+ Project (KA+ 2018-1-IT02-KA203-048519), which involved an international partnership led by the University of Naples Federico II. ALEAS is devoted to bachelor students enrolled in non-STEM degree programs. It encompasses a knowledge structure for the introductory statistics course organized into a hierarchical structure defined by Areas, Topics, and Units. Each learning Unit considers three Dublin descriptors to assess and improve students' ability in Statistics (a brief overview of the Dublin descriptors is provided in Chapter 3). The tutoring system implemented in ALEAS combines two approaches: for each Topic, a multidimensional latent class IRT model (Bartolucci, 2007) estimates items' difficulty and discrimination power and simultaneously categorizes students into the latent classes; for each Area, average ability levels were computed according to the estimated topic-level latent class IRT models, and then an archetypal analysis (Cutler and Breiman, 1994) clustered students into homogeneous groups. Based on students' categorization, the system provides tailored feedback suggesting students either to continue training in the current Area or to progress to a new one. ALEAS also includes an animated tutoring agent, 3D cartoons, and vignettes to facilitate learning some essential statistical topics and engage students in their progressive achievement. For more details about the ALEAS architecture, methodology, and preliminary results, see Adabbo et al. (2021) and Pacella et al. (2022).

Chapter 3

A multidimensional approach for assessing students' ability

Students' ability is a complex theoretical concept, whose evaluation requires disentangling between its different dimensions. To this end, the main challenges regard the construction of the knowledge domain to investigate, the definition of what dimensions qualify the considered ability, and the development of appropriate indicators for their measurement. In addition, a comprehensive assessment of students' proficiency should also consider individual characteristics affecting students' achievements and performance (e.g., emotional, psychological, and motivational aspects).

Herein, a multidimensional approach based on the Dublin descriptors (Gudeva et al., 2012) is proposed. Accordingly, sets of multiple-choice

The content of this chapter is included in the paper:
Fabbricatore, R., Bakk, Z., Di Mari, R., de Rooij, M., & Palumbo, F. (2022). A non-parametric multilevel latent variable model for handling students' learning activities data. *Submitted.*

questions were developed to measure the considered dimensions of students' ability. As detailed below, each question distinguishes between wrong, correct, and partially correct answers, allowing to detect various levels of gained knowledge. Moreover, the proposed students' ability assessment involves different topics over time and a set of psychological characteristics related to students' performance. The practical application specifically refers to the context of learning Statistics in non-STEM degree courses; however, the proposed approach can be generalized to any different knowledge domain.

The chapter is organized as follows: Section 3.1 briefly presents the application context, namely learning Statistics in non-STEM degree courses; Section 3.2 provides a detailed description of data collection, participants, and adopted measures; Section 3.3 includes descriptive statistics regarding students' performance and cognitive and psychological characteristics considered in the study. Note that the collected data are used for the empirical application of the two proposed statistical approaches presented in Chapter 4 and Chapter 5, respectively.

3.1 Learning Statistics in non-STEM degree courses

Nowadays, statistical knowledge has become part of the new generation's knowledge base (Istat, 2022). Indeed, the great availability of data in the modern age brought the need for proper data analysis tools, lending Statistics a fundamental role in most Higher Education curricula. In STEM and non-STEM degree courses, Statistics facilitates students' ability to understand and communicate the data appropriately, fostering world interpretability. Nevertheless, Statistics not only favors hard skills development; it also improves soft skills, such as critical thinking and quantitative reasoning (Lehman and Nisbett, 1990; Ben-Zvi and Makar, 2016), which are essential in daily life.

Despite its relevance, Statistical literacy is still challenging, especially for students enrolled in Human and Social degree courses. Such students often believe that Statistics is not essential for their degree programs and careers. Thus, they are less prone to study quantitative subjects and consider Statistics a burden (Sesé et al., 2015). In addition to the lack of motivation, often such students feel unconfident, discouraged, and anxious during their statistical learning experience, resulting in low performance and poor achievements in Statistics (Walker and Brakke, 2017).

The anxiety in students affording Statistics (statistical anxiety) negatively affects: *(i)* the learning experience, sometimes causing drop-out during the course, *(ii)* the exam preparation phase, reducing cognitive resources, and *(iii)* academic outcomes such as failing the exam or getting lower grades (Siew et al., 2019; Macher et al., 2013; Eysenck et al., 2007). Anxiety occurring during test situations (test anxiety) also influences academic performance due to feelings of worry, tension, and over-arousal (Rajiah et al., 2014; Fabbricatore et al., 2022b). Among psychological factors related to students' performance in Statistics, attitudes toward Statistics and self-efficacy stand out, influencing their willingness to participate in statistical courses, enjoyment of statistical learning, use of statistical knowledge for real-life problems, and learning motivation (Judi et al., 2011; Rosli and Maat, 2017; Richardson et al., 2012; Honicke and Broadbent, 2016).

Other relevant dispositional antecedents relating to academic life are academic motivation, academic procrastination, self-regulation, and the use of cognitive strategy. Indeed, frustration due to lack of motivation and the tendency to delay academic tasks have a detrimental effect on students' achievement and performance (Cokley et al., 2001; Steel et al., 2001). On the other hand, the ability to manage the learning process and the use of suitable cognitive strategy, e.g., integrating notions acquired during lessons with other course materials or repeating important topics several times to remember them better, improve the knowledge outcomes. Accordingly, students' engagement in the educational process enhances performance in Statistics, increasing their interest, participation, and study regularly (Budé et al., 2007; Lavidas et al., 2020).

Finally, cognitive factors, such as math knowledge, are also assumed to have a prominent role in determining students' performance in Statistics (Johnson and Kuennen, 2006). Indeed, several studies highlighted that severe math shortcomings could constitute a roadblock to complete statistics courses successfully (Johnson and Kuennen, 2006; Rabin et al., 2018).

Given the importance of these factors, accounting for them during the analysis of students' performance provides a deeper understanding of the learning process, and consequently allows the development of more tailored feedback in a recommender system.

3.2 Data collection

This section describes the data at hand. Firstly, the participants and the ability assessment procedure are illustrated. An overview of the developed and adopted measures follows. Finally, descriptive statistics about students' performance in Statistics and the considered psychological variables are provided. Regarding the latter, according to the above-discussed literature, the following cognitive and psychological characteristics were evaluated: math knowledge, statistical anxiety, attitude toward Statistics, self-efficacy, test anxiety, cognitive strategies, self-regulation, academic procrastination, academic motivation, and engagement in Statistics.

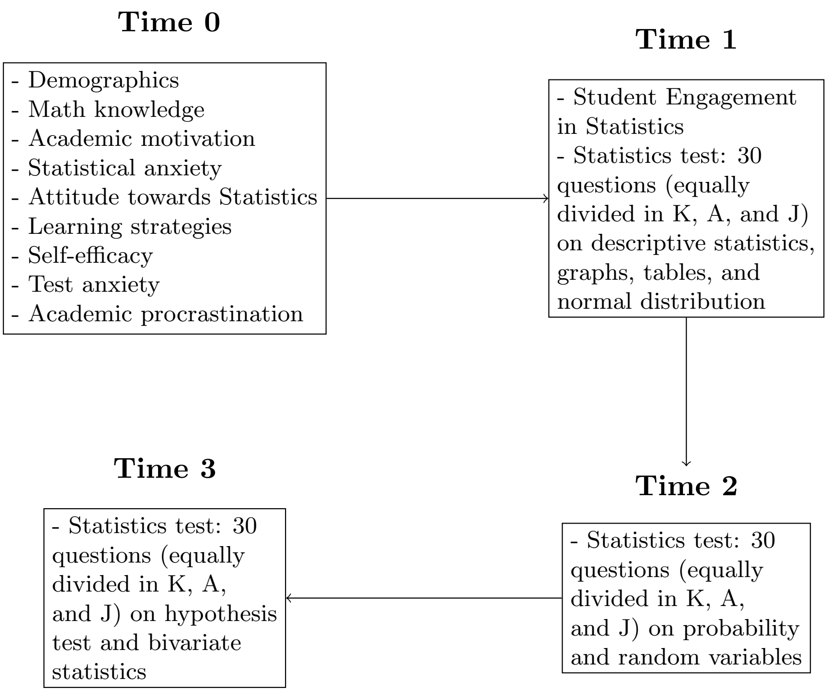
3.2.1 Participants and Procedure

The study involved $n = 202$ Italian students enrolled in the first year of the degree in Psychology at the University of Naples Federico II, attending the introductory Statistics course. Participants were predominantly female (83.6%) with age ranged between 18 and 43 (mean = 19.7, sd = 2.77). At the beginning of the course, students were invited to fill out an online self-report questionnaire including socio-demographic questions and psychometric scales assessing psychological variables related to learning Statistics.

Throughout the course, students were encouraged to use the Moodle platform (Gamage et al., 2022) to practice and monitor their progress in Statistics. In addition, at the end of each main statistical module, students were asked to take a summary test to check their acquired knowledge. The analyzed data comprise students' responses to these summary Statistics tests. In particular, the knowledge structure was organized into three main areas, each assessed at a different time point: descriptive statistics, graphs, tables, and Gaussian distribution (time 1), probability and random variables (time 2), and hypothesis testing and bivariate statistics (time 3). For each learning area, students' ability was measured in a multidimensional way. The following section describes the considered dimensions, herein named Knowledge (K), Application (A), and Judgment (J). Figure 3.1 illustrates the data collection flow. Notice that the entire data collection was carried out through the Moodle platform, one of the most widely used learning management systems for delivering courses and learning material, assessing students' competencies, and providing feedback in higher education (Gamage et al., 2022). During data collection, some students dropped out. More specifically, $n = 166$ students remained at time 2, and $n = 126$ at time 3.

The study was approved by the Ethical Committee of Psychological Research of the University of Naples Federico II (protocol number 26/2022).

Figure 3.1: Data collection flow: the four considered time points.



Note: K, A, and J refer to the Knowledge, Application and Judgment ability's dimensions, respectively.

3.2.2 Measures

Statistics performance

Students' ability was assessed in a multidimensional way. The so-called Dublin descriptors[1] were used for this aim, representing general statements used to evaluate the knowledge depth a student has achieved within a specific topic at the end of each cycle of higher education studies (Gudeva et al., 2012). More specifically, they define the following five learning objectives:

- *Knowledge and understanding*: the ability to demonstrate knowledge and understanding, including a theoretical, practical and critical perspective on the topic;
- *Applying knowledge and understanding*: the ability to apply the knowledge identifying, analyzing and solving problems sustaining an argument;
- *Making judgments*: the ability to gather, evaluate and present information exercising appropriate judgment;
- *Communications skills*: the ability to communicate ideas, problems and solutions effectively and disseminate them to a non-specialist audience;
- *Learning skills*: the ability to identify learning needs and fill the knowledge gaps.

Since data collection was performed exploiting a technological platform, only the first three descriptors are considered. As stated before, these three

[1] https://aec-music.eu/userfiles/File/Framework_for_Qualifications_of_European_HE_Area.pdf

ability dimensions are herein named, for the sake of simplicity, Knowledge (K), Application (A), and Judgment (J), respectively.

In particular, each Statistics test consists of 30 multiple-choice questions, equally divided into K, A, and J, each question having four answer options and three different response scores: two credits for totally correct answers; one credit for partially correct answers, and no credit for wrong answers. Blank responses are considered missing values. Referring to the hypothesis testing domain, Figure 3.2 provides an example of questions, each one relating to one of the three considered Dublin descriptors.

The R package `exams` (Grün and Zeileis, 2009) was used to prepare the test[2]. Interested readers can refer to Fabbricatore et al. (2021) for more details.

Math knowledge and psychological factors

Math knowledge and psychological factors were assessed through validated psychometric scales. In particular, the *Mathematical Prerequisites for Psychometrics* (PMP; Galli et al., 2008) was used to evaluate the basic mathematics abilities usually required for an introductory Statistics course in a Psychology degree program. The scale consists of 30 multiple-choice questions with only one correct answer covering 6 knowledge domains: Fractions, Operations, Set theory, Equations, Relations, and Probability. Students gained one point for each correct response, whereas wrong and missing answers received no credits. Thus, scores ranged between 0 and 30.

Students' statistical anxiety was assessed through the *Statistical Anxiety Scale* (SAS; Chiesi et al., 2011), a 24-item scale embracing three different di-

[2]Many of the questions were developed during the "Adaptive LEArning in Statistics" (ALEAS) Erasmus+ project (KA+2018-1-IT02-KA203-048519).

mensions of the statistical anxiety: *examination anxiety*, referring to anxiety students encounter while attending statistics classes or taking statistics tests; *interpretation anxiety*, occurring when students are required to interpret or make a decision about statistical data; *fear of asking for help*, concerning the anxiety experienced when requesting the help of a peer, a tutor, or a professor in understanding particular topics. For each item, responses are on a 5-point Likert scale, from 1 (no anxiety) to 5 (very strong anxiety). The score of each sub-scale is obtained by summing up the score observed on the corresponding items, resulting in a theoretical range from 8 to 40.

A multidimensional evaluation of students' attitude towards statistics was obtained through the *Survey of Attitudes toward Statistics* (SATS; Chiesi and Primi, 2009). Specifically, the instrument contains 28 items assessing the following attitude components: affection (students' positive and negative feelings concerning Statistics), cognitive competence (students' belief about their ability in Statistics), value (perceived usefulness and relevance of Statistics in their personal and professional life), difficulty (the consideration of Statistics as a difficult subject). Students were required to express their degree of agreement for each item on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The score of each dimension is computed as the sum of the scores observed on the corresponding items. Thus, the theoretical sub-scale range is equal to 6-42 for affect and cognitive competence, 9-63 for value, and 7-49 for difficulty.

Self-efficacy, test anxiety, cognitive strategies and self-regulation were assessed using *The Motivated Strategies for Learning Questionnaire* (MSLQ; Bonanomi et al., 2018). The Self-efficacy sub-scale comprises 9 items referring to expectancy for success in a given knowledge domain, perceived ability to complete a task, and self-confidence. Test anxiety sub-scale consists of 4 items measuring students' anxiety and concern about taking exams. Cognitive strategies evaluation is based on 12 items focused on some suitable strategies to use to obtain a good achievement in a learning path. On the other hand, the application of useful cognitive and metacognitive strategies to effectively supervise the learning process was measured through the self-regulation sub-scale that includes 10 items. For all the considered sub-scales, students were asked to indicate their degree of agreement on a 5-point Likert scale ranging from 1 (not at all true for me) to 5 (very true for me). The theoretical sub-scale range, provided by the sum of the score observed on the corresponding items, is equal to 9-45 for self-efficacy, 4-20 for test anxiety, 12-60 for cognitive strategy, and 10-50 for self-regulation.

Academic procrastination was evaluated employing *The Academic Procrastination Scale - Short Form* (APS-SF; Yockey, 2016; Fabbricatore et al., 2022a), a scale consisting of 5 items with a 5-point Likert response scale from 1 (I disagree) to 5 (I agree). The total score is defined as the arithmetic mean of the score observed on the single items (theoretical range 1-5), where higher scores indicate a greater tendency to delay academic tasks despite the negative outcomes it causes (e.g., low performance, distress, anxiety).

Academic motivation was assessed using *The Academic Motivation Scale* (AMS; Alivernini and Lucidi, 2008) that considers 5 dimensions of motivation, each evaluated through 4 items. In particular, students were required to indicate their degree of agreement to possible responses to the question "Why do you go to University?" on a 7-point Likert scale, from 1 (does not correspond at all) to 7 (corresponds exactly). Herein, the overall indicator of students' motivational orientation was considered, named Relative Autonomy Index (RAI), which is computed as a weighted sum of the academic motivation dimensions. In particular, the following expression is used:

$$\text{RAI} = 2 \cdot (\text{Intrinsic Motivation}) + 1 \cdot (\text{Identified Regulation}) - 1 \cdot (\text{External Regulation}) - 2 \cdot (\text{Amotivation}) + 0 \cdot (\text{Introjected Motivation}).$$

Accordingly, RAI theoretically ranges from -72 to 72, where positive scores indicate more autonomous regulation, whereas negative scores reveal more controlling regulation.

Student engagement in Statistics was measured by *The scale of Student Engagement in Statistics* (SSE-S, Whitney et al., 2019), grounded on a three-factor structure: affective (8 items), behavioral (9 items), and cognitive (7 items) engagement. The first dimension considers interest, curiosity, and enjoyment in learning statistics. The second one concerns learners' observable behaviors related to study Statistics (e.g., regular study, participation during lessons, interaction with teachers). The third aspect refers to the application of cognitive strategies to connect, reexamine, and assess own learning of Statistics. For all the considered sub-scales, students were asked to indicate their degree of agreement on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The total score for each sub-scale is obtained by averaging the score observed on the related items, resulting in a theoretical range from 1 to 5.

3.3 Descriptive statistics

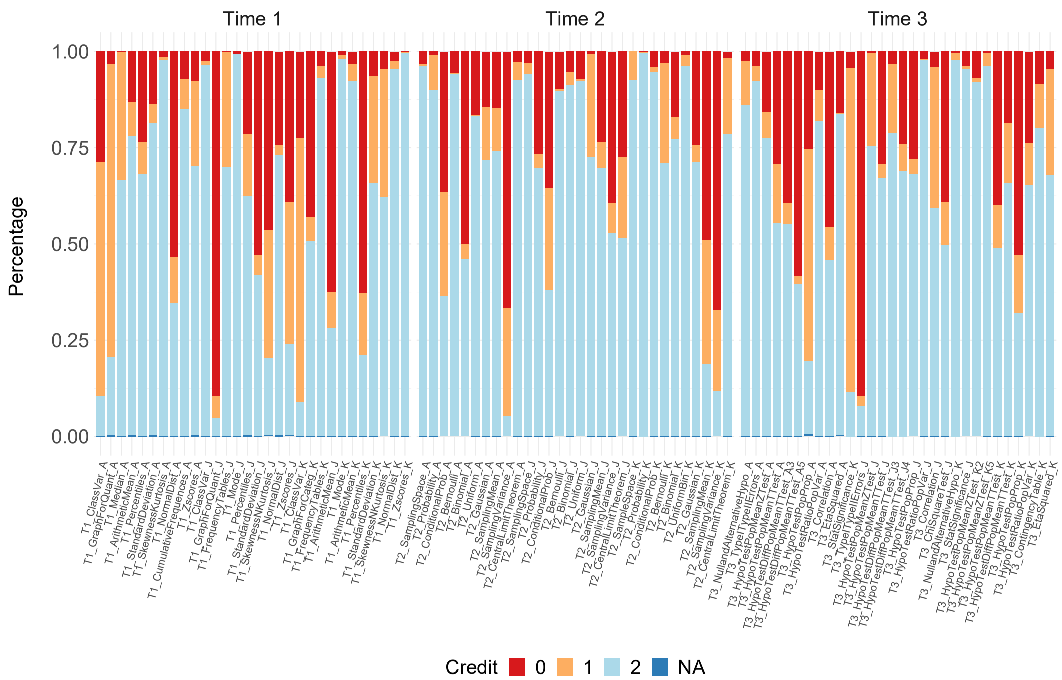
Descriptive statistics about students' performance in Statistics are reported in Figure 3.3 while Table 3.1 summarizes the related cognitive and psychological factors.

Figure 3.2: Question example for each of the considered Dublin descriptors.

Knowledge	**Question** Which of the following statements related to the parametric test for hypothesis testing is true? **Answerlist** a) The acceptance and rejection regions are subsets of the parametric space b) The acceptance and rejection regions are compatible subsets of the parametric space c) The acceptance and rejection regions can be partially overlapped d) The acceptance and rejection regions belong to two different parametric spaces
Application	**Question** A psychologist expert in specific learning disabilities (SLD) wants to verify whether children with SLD have a normal overall intellectual functioning as reported by the DSM-5. Thus, the psychologist administered the WISC-IV scale to 38 children with SLD to assess their Intelligence Quotient (IQ). The results showed an average score of 95.2. Considering that the IQ mean for the reference population is 100 with a standard deviation of 15, calculate the value of the statistic test to test the hypothesis $H_0 : \mu = 100$ versus $H_1 : \mu < 100$. **Answerlist** a) -1.22 b) 1.22 c) -6.01 d) -0.2
Judgment	**Question** The z test for hypothesis testing on the mean can be used to investigate: **Answerlist** a) if a group of patients has average cholesterol levels b) if the attention levels are higher in individuals who sleep regularly than in individuals who do not sleep regularly c) if there is an association between the nationality of the individuals and the favorite car brand d) the amount of banknotes produced in Italy in one year

Note: The letter a) corresponds to the totally correct answer, and the letter b) to the partially correct one.

Figure 3.3: Distribution of the answers for each question according to the three time points.



Note: Question labels are divided in three parts providing specific information: the time (e.g., "T1_"), the topic (e.g., "Median"), and the Dublin descriptor dimension (e.g., "_A"). Note that no credit was given to wrong answers, one credit to partially correct answers, and two credits to totally correct answers; missing responses were considered missing values (NA). Note also that the dropouts at time 2 and time 3 were not considered in the depicted response distributions; thus, the percentages on the y-axis refer to a different number of total observations: 202 for time 1, 166 for time 2, and 126 for time 3.

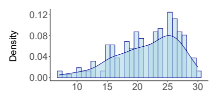
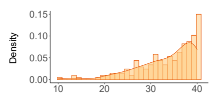
As regards Statistics performance, response distributions show that students at times 2 and 3 scored well on a slightly larger number of items than at time 1. In particular, 20 questions about descriptive statistics obtained more than 50% of fully correct answers compared to the 24 questions on probability and random variables, and the 22 questions on hypothesis testing and bivariate statistics.

The multidimensional evaluation through the Dublin descriptors allows also detecting impairments for specific competencies within each topic. For example, students proved to master the theoretical definition of the arithmetic mean and how to apply it to solve exercises, but not to have enough competence for making an appropriate judgment given a set of available information. At times 2 and 3, a greater balance in the students' ability dimensions is overall observed, with some specific topics reporting differences in the response distribution for the considered dimensions. For example, although students acquired the theoretical definition of conditional probability at time 2, they still experienced difficulties during problem-solving and judgment exercises involving reasoning about conditional probability. Another example from hypothesis testing highlights students' difficulty in providing a proper judgment about the type I and II errors, even if they reported good scores in the corresponding application question. Moreover, as concerns the hypothesis test on the population mean, results show that the mastery of the Student's t-test was more challenging for students than the z-test, especially for the Knowledge dimension. Finally, note that missing responses computed on the valid cases are less than 5% for each question.

Regarding cognitive and psychological factors related to learning Statistics, descriptive analyses show that students have a medium-high level of math knowledge (more than 21 out of 30 correct answers, on average).

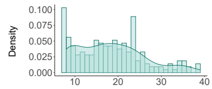
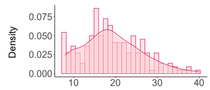
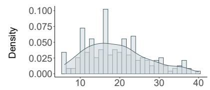
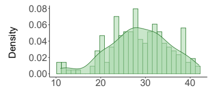
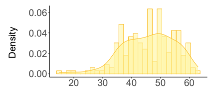
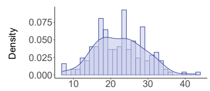
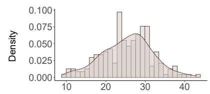
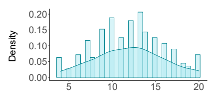
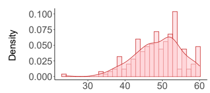
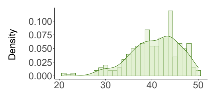
Examinations represented their main source of statistical anxiety, with a lower level of anxiety encountered when interpreting statistical data and asking for help. General test anxiety appears to be moderate, revealing a lower impact on students' feelings than the more specific statistical anxiety. Data reveal that students' feelings concerning Statistics are more negative than positive and that Statistics is considered a subject difficult to study even if useful and relevant for the personal and professional life. Students' self-confidence about the ability to learn Statistics (cognitive competence) is quite high on average. Accordingly, scores related to the use of cognitive and metacognitive strategies to supervise the learning process are also high, and the distributions are negatively skewed. On the other hand, on average, moderate levels of self-efficacy and engagement in Statistics emerge. Taking into account the engagement in Statistics, the value of the behavioral component (such as studying regularly and participating during lessons) is slightly greater than affective and cognitive values. Finally, students report a high level of autonomous academic motivation and a low tendency to delay academic tasks intentionally (academic procrastination).

Table 3.1: Descriptive statistics for the considered cognitive and psychological covariates.

Scale	Dimension	Theoretical range	Mean	Sd	Histogram and density plot
PMP		0-30	21.58	5.13	
SAS	Examination	8-40	33.61	6.16	

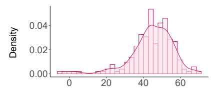
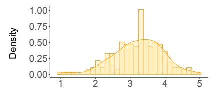
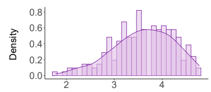
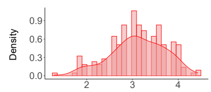
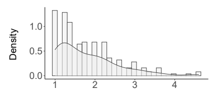
Continued on next page

Table 3.1: Descriptive statistics for the considered cognitive and psychological covariates (continued).

Scale	Dimension	Theoretical range	Mean	Sd	Histogram and density plot
	Asking for help	8-40	18.33	7.82	
	Interpretation	8-40	19.17	7.03	
SATS	Affect	6-42	19.09	7.83	
	Cognitive competence	6-42	28.55	6.80	
	Value	9-63	45.94	9.00	
	Difficulty	7-49	22.37	6.54	
MSLQ	Self-efficacy	9-45	25.41	6.37	
	Test anxiety	4-20	11.96	3.94	
	Cognitive strategy	12-60	48.83	6.31	
	Self-regulation	10-50	40.59	5.32	

Continued on next page

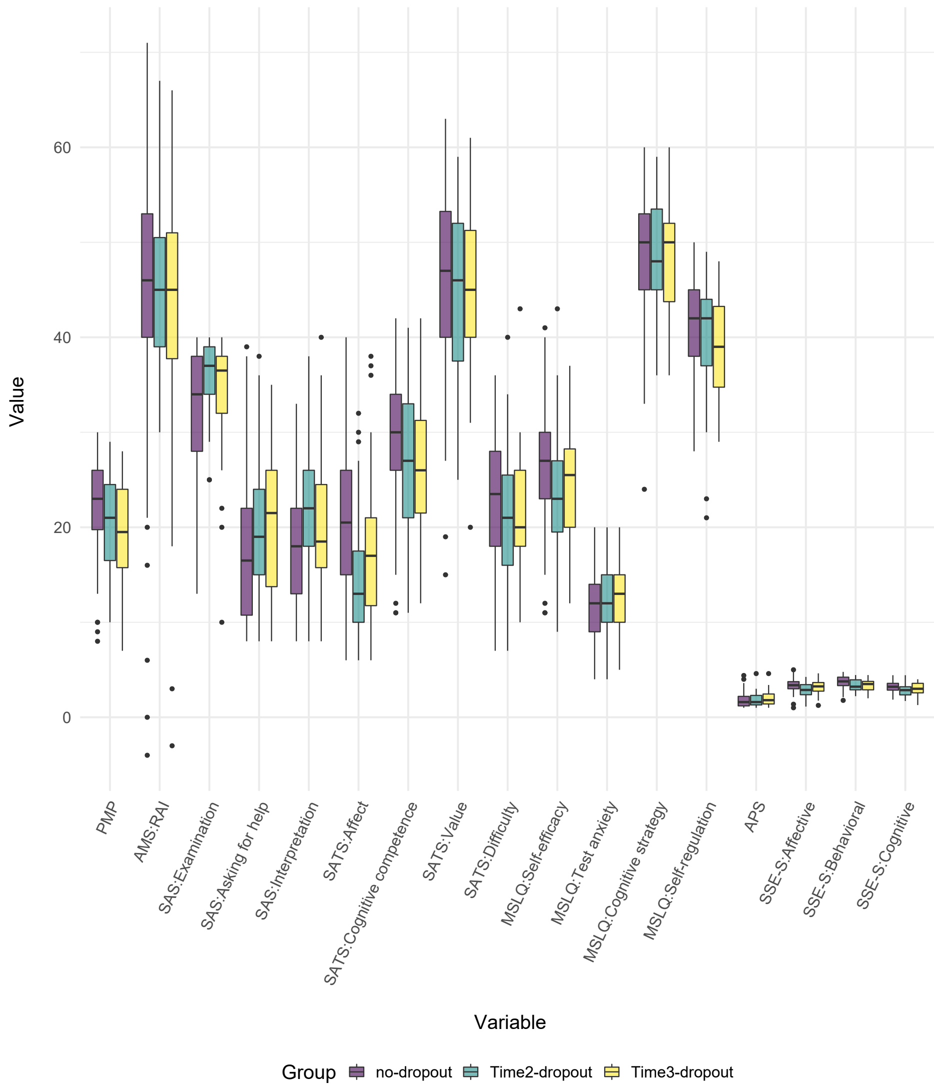
Table 3.1: Descriptive statistics for the considered cognitive and psychological covariates (continued).

Scale	Dimension	Theoretical range	Mean	Sd	Histogram and density plot
AMS	RAI	-72 - 72	44.53	11.88	
SSE-S	Affective	1-5	3.23	0.70	
	Behavioral	1-5	3.60	0.63	
	Cognitive	1-5	3.11	0.60	
APS		1-5	1.81	0.76	

Note: PMP = Mathematical Prerequisites for Psychometrics; SAS = Statistical Anxiety Scale; SATS = Survey of Attitudes toward Statistics; MSLQ = Motivated Strategies for Learning Questionnaire; AMS = Academic Motivation Scale; APS = Academic Procrastination Scale; SSE-S = Scale of Student Engagement in Statistics; RAI = Relative Autonomy Index.

To characterize students dropping out during learning, descriptive statistics about psychological variables were computed according to three dropout groups: no-dropout (students involved until the end of the course), Time2-dropout (students dropping out at time 2), Time3-dropout (students dropping out at time 3). The corresponding boxplots are depicted in Figure 3.4. As can be seen, compared to the no-dropout group, students who dropped out during the learning process reported a poorer mathematics background, a more negative attitude toward Statistics (especially for the affect and cognitive competence dimensions), and a lower level of self-efficacy and engagement in Statistics. Regarding the remainder variables, there were no noticeable differences between the no-dropout group and students dropping out. Time2- and Time3-dropout groups presented remarkable differences only for some variables, such as examination and interpretation components of statistical anxiety, with lower scores for students who dropped out at time 3. Conversely, the latter presented higher scores for the affective dimension of both attitude and engagement, highlighting the relevance of the affective component in buffering students' dropout. Concerning sex differences, results showed the following percentage frequencies according to the dropout groups: the 70% of males and 61% of females did not drop out during learning (no-dropout), 12% of males and 19% of females dropped out at time 2, 18% of males and 20% of females dropped out at time 2 (Time3-dropout).

Figure 3.4: Boxplots of the psychological variables scores according to the dropout groups.



Chapter 4

A multilevel approach for multidimensional latent variables

This chapter introduces a novel strategy exploiting non-parametric multilevel latent variable models (Vermunt, 2003) to cluster students into homogeneous groups according to their ability level.

Previous studies with a similar aim (see, e.g., Fagginger Auer et al., 2016; Vermunt, 2003) focus on *unidimensional* specifications. As a relevant methodological contribution, this proposal specifies a multidimensional latent structure at a low level. The latter, conditional on Level 2 class membership, is assumed to be composed of independent Level 1 discrete latent class variables, each measuring a distinct learning dimension.

The content of this chapter is included in the paper:
Fabbricatore, R., Bakk, Z., Di Mari, R., de Rooij, M., & Palumbo, F. (2022). A non-parametric multilevel latent variable model for handling students' learning activities data. *Submitted.*

Data at hand consist of answers to items that cover different Statistics–related topics and ability dimensions collected at several time points. For a detailed description, see Chapter 3. As shown below, it is possible to recast the resulting multilevel multi-dimension structure to the typical structure of multilevel latent class models using a suitable rearrangement of the data layout. In particular, time–points can be viewed as low–level/Level 1 unit, clustered within individuals high–level/Level 2. Moreover, demographic and several other psychological variables can be included in the model as (high-level) predictors of students' performance.

4.1 Multilevel Latent Class model

Let $\left(Y_{sd1}^{(t)}, \ldots, Y_{sdk_t}^{(t)}\right)'$ be the vector of responses for individual $s = 1, \ldots, N$ on the k_t indicators at time point $t = 1, \ldots, T$, with $\sum_t k_t = K$. The considered model was designed to deal with $D = 3$ skills that are defined according to the Dublin descriptors (Knowledge, Application, and Judgment). Note that the model can be generalized to any different number of skills and any different definitions. Moreover, let $\mathbf{c}_s$ be a vector of (level 2) covariates (demographic and psychological) for the s-th student, which will be exploited as a predictor of higher-level class membership.

This type of data can be re-arranged in the form of one record per time point, with a unique id linking the responses of each sample unit over time. The proposed re-arrangement of the data can be recognized as a nested/hierarchical structure consisting of D *versions* of the considered items nested within time points and three-time points nested within persons. Therefore, multilevel latent class analysis can be used to handle this data structure to address the research aim. In the more general framework of multilevel latent variable models, hierarchical data are handled by an analytic structure exploiting discrete and/or continuous latent variables at each level (see, e.g., Vermunt, 2003). The non-parametric approach was adopted for its flexibility.

In particular, the considered latent structure is specified as follows. At the lower level, D discrete latent variables correspond to the considered dimensions (herein Knowledge, Application, and Judgment). Each of these D dimensions is measured through a distinct set of multiple-choice questions, coded as ordinal indicators, with q categories, indicated with $j = 0, 1, \ldots, J$. At the higher level, another discrete latent variable makes it possible to cluster students based on their likelihood to be in one of the lower–level classes for each Dublin descriptor.

Technically, let $X_{sd}^{(t)}$ be the d-th categorical latent variable defined at a lower level (Level 1) for the t-th time point, taking value $i = 1, \ldots, I$. The total number of latent classes I is assumed to be equal for the D latent variables at Level 1. Similarly, W_s denotes the discrete latent variable defined at Level 2, taking value $m = \{1, \ldots, M\}$. Note that, conditioning on W_s, $X_{sd}^{(t)}$ and $X_{sf}^{(r)}$ are independent, for all $d \neq f$ and $t \neq r$.

Therefore, the multilevel latent class model specifies the joint probability of the observed response vector $\mathbf{Y}_s = \left(Y_{s11}^{1}, \ldots, Y_{s3k_T}^{(T)}\right)'$, given the available covariates, as follows

$$P\left(\mathbf{Y}_s|\mathbf{c}_s\right) = \sum_{m=1}^{M} P\left(W_s = m|\mathbf{c}_s\right) \prod_{t=1}^{T} \prod_{d=1}^{D} \left\{ \sum_{i=1}^{I} P\left(X_{sd}^{(t)} = i|W_s = m\right) P\left(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i\right) \right\}. \tag{4.1}$$

The path diagram for the model (4.1) is displayed in Figure 4.1.

The joint conditional response probability, because of the local independence assumption, can be factorized as

$$P(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i) = \prod_{h=1}^{k_t} P(Y_{sdh}^{(t)}|X_{sd}^{(t)} = i).$$

As for ordinal items, the conditional response probabilities can be parametrized by means of the following adjacent-category ordinal logits

$$P(Y_{sdk}^{(t)} = j|X_{sd}^{(t)} = i) = \frac{\exp\left(\sum_{h=2}^{J} \beta_{j|i}^{(t)}\right)}{1 + \sum_{l=2}^{J} \exp\left(\sum_{q=2}^{l} \beta_{q|i}^{(t)}\right)}, \tag{4.2}$$

with $j = 2, \ldots, J$. The adjacent-category ordinal logit model is typically used to model the ordered response when partial credit is allowed during an evaluation process, as in the proposed students' ability assessment (Bartolucci et al., 2015; Eggert and Bögeholz, 2010). In particular, partial credit indicates intermediate performance levels on an item, moving beyond the binary right/wrong answer coding rationale.

Let $\mathbf{z}$ be a vector of $T - 1$ time dummies - the dummy for $t = 1$ is excluded for redundancy - with generic element $\mathrm{z}_{(t-1)}$, where $t = 2, \ldots, T$. Low level conditional class membership probabilities can be parameterized by means of the following multinomial logistic regressions

$$P(X_{sd}^{(t)} = i|W_s = m) = \frac{\exp\left(\gamma_{0im} + \gamma_i^{(t)} \mathrm{z}_{(t-1)}\right)}{1 + \sum_{r=2}^{I} \exp\left(\gamma_{0rm} + \gamma_r^{(t)} \mathrm{z}_{(t-1)}\right)}, \tag{4.3}$$

where γ_{0im} is a random intercept, with $i = 2, \ldots, I$, and $m = 1, \ldots, M$, and $\gamma_i^{(t)}$ is a time effect, with $t = 2, \ldots, T$. Note that, for $t = 1$, model (4.3) reduces to a standard random-intercept-only multinomial logistic regression.

On a practical level, the high-level class membership independent time coefficients and the random intercepts of Equation (4.3) can be conveniently handled by considering an *expanded* design matrix $\mathbf{D}$ of length $M \times (n \times T)$, with generic row equal to $(\tilde{\mathbf{d}}_m, \mathbf{z})'$. As an example, the first entry of $\mathbf{D}$, i.e. $\tilde{\mathbf{d}}_1$, is equal to 1, and the remaining $M-1$ are equal to 0 for the first $n \times T$ units, and so on.

Analogously, for W_s the multinomial logistic model can be considered

$$P\big(W_s = m|\mathbf{c}_s\big) = \frac{\exp\big(\delta_{0m} + \delta_m \mathbf{c}_s\big)}{1 + \sum_{r=2}^{M} \exp\big(\delta_{0r} + \delta_r \mathbf{c}_s\big)},$$

where δ_{0m}, and δ_m are the intercept, and the covariate effects, respectively, with $m = 2, \ldots, M$.

The model (4.1) is identified, in a *generic sense* (see Allman et al., 2009), provided that $P\big(Y_{sdh}^{(t)}|X_{sd}^{(t)} = i\big) \neq P\big(Y_{sdh}^{(t)}|X_{sd}^{(t)} = j\big)$ for all $i \neq j$, and $\gamma_{0im} \neq \gamma_{0in}$ for all $m \neq n$ (Di Mari et al., 2022).

4.1.1 Class prediction

At population level, it is assumed that each student belongs to only one higher level latent class W_s. Equally, *true* lower level class membership is *crisp* (or hard), yet unobserved.

Given a full response pattern $\mathbf{Y}_s$ and a vector of covariates $\mathbf{c}_s$, Level 2 class membership can be predicted, using the Bayes' theorem, by means of the following posterior probabilities

$$P\big(W_s|\mathbf{Y}_s, \mathbf{c}_s\big) = \\ = \frac{P\big(W_s = m|\mathbf{c}_s\big) \prod_{t=1}^{T} \prod_{d=1}^{D} \Big\{\sum_{i=1}^{I} P\big(X_{sd}^{(t)} = i|W_s = m\big) P\big(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i\big)\Big\}}{\sum_{h=1}^{M} P\big(W_s = h|\mathbf{c}_s\big) \prod_{t=1}^{T} \prod_{d=1}^{D} \Big\{\sum_{i=1}^{I} P\big(X_{sd}^{(t)} = i|W_s = h\big) P\big(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i\big)\Big\}}. \tag{4.4}$$

Similarly, the posterior probability that student s belongs to class i for the d-th lower level latent class variable at time t can be obtained as

$$P\big(X_{sd}^{(t)} = i|\mathbf{Y}_{sd}, \mathbf{c}_s\big) = \sum_{m=1}^{M} P\big(W_s|\mathbf{Y}_s, \mathbf{c}_s\big) \frac{P\big(X_{sd}^{(t)} = i|W_s = m\big)P\big(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i\big)}{P\big(\mathbf{Y}_{sd}^{(t)}\big)}. \tag{4.5}$$

These quantities are readily available for the observed data from the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) used for model fitting. Equations (4.4) and (4.5) can be used also to predict class membership, given parameter estimates, for *out-of-sample* units. This is done by simply plugging in the new data, and evaluating both quantities at the ML estimates. Subsequently, well-known classification rules, like modal or proportional assignments, can be used to cluster high/low level units.

4.1.2 Parameter estimation

By letting θ be the vector collecting all model parameters, the log-likelihood function for model (4.1) can be written as follows

$$\ell(\theta) = \sum_{s=1}^{N} \log P\big(\mathbf{Y}_s|\mathbf{c}_s\big),$$

which has to be maximized with respect to θ in order to obtain the maximum likelihood (ML) estimates $\widehat{\theta}$. Such estimates can be computed either by direct maximization, or exploiting the EM algorithm (Dempster et al., 1977). To obtain the best of both worlds, hybrid techniques are also available. These exploit the EM algorithm in earlier stages of model fit - i.e. far from the (local) optimum - then switching to Newton-Raphson for faster convergence when closer to the (local) maximum. Computations have been performed in Latent GOLD 6.0 (Vermunt and Magidson, 2020), which implements the above mixed strategy[1].

[1]The Latent GOLD syntax for the current application is provided in Appendix A

In what follows, the EM algorithm to compute the ML estimate $\widehat{\theta}$ is outlined, with specific emphasis to the E-step. This extends the EM presented by Vermunt (2003, 2008) for multilevel LCA. Adopting the classical EM terminology, let

$$u_{sm} = \begin{cases} 1, & \text{if } W_s = m \\ 0, & \text{otherwise.} \end{cases}, \quad v_{sdim}^{(t)} = \begin{cases} 1, & \text{if } X_{sd}^{(t)} = i, \quad W_s = m, \\ 0, & \text{otherwise.} \end{cases}$$

be the (unobserved/missing) *augmenting variables*, where $v_{sdim}^{(t)}$ is defined for all combinations of $d = 1, \ldots, D$, and $t = 1, \ldots, T$. Given both observed and unobserved data, the *complete data log-likelihood* CDLL(θ) can be formulated as

$$\begin{aligned} \text{CDLL}(\theta) = & \sum_{s=1}^{N} \sum_{m=1}^{M} u_{sm} \log \left\{ P\big(W_s | \mathbf{c}_s\big) \right\} + \\ & \sum_{s=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{i=1}^{I} v_{sdim}^{(t)} \log \left\{ P\big(X_{sd}^{(t)} = i | W_s = m\big) \right\} + \\ & \sum_{s=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{i=1}^{I} v_{sdim}^{(t)} \log \left\{ P\big(\mathbf{Y}_{sd}^{(t)} | X_{sd}^{(t)}\big) \right\}. \end{aligned}$$

In the E-step, the missing data u_{sm} and $v_{sdim}^{(t)}$ are replaced by conditional expectations to compute the following expected CDLL

$$\begin{aligned} \mathbb{E}\left[\text{CDLL}(\theta)\right] = & \sum_{s=1}^{N} \sum_{m=1}^{M} \widehat{u}_{sm} \log \left\{ P\big(W_s | \mathbf{c}_s\big) \right\} + \\ & \sum_{s=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{i=1}^{I} \widehat{v}_{sdim}^{(t)} \log \left\{ P\big(X_{sd}^{(t)} = i | W_s = m\big) \right\} + \\ & \sum_{s=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{i=1}^{I} \widehat{v}_{sdim}^{(t)} \log \left\{ P\big(\mathbf{Y}_{sd}^{(t)} | X_{sd}^{(t)}\big) \right\}, \end{aligned}$$

where the desired quantities are computed respectively as

$$\widehat{u}_{sm} = \frac{P(W_s = m|\mathbf{c}_s) \prod_{t=1}^{T} \prod_{d=1}^{D} \left\{ \sum_{i=1}^{I} P(X_{sd}^{(t)} = i|W_s = m) P(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i) \right\}}{\sum_{h=1}^{M} P(W_s = h|\mathbf{c}_s) \prod_{t=1}^{T} \prod_{d=1}^{D} \left\{ \sum_{i=1}^{I} P(X_{sd}^{(t)} = i|W_s = h) P(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i) \right\}},$$

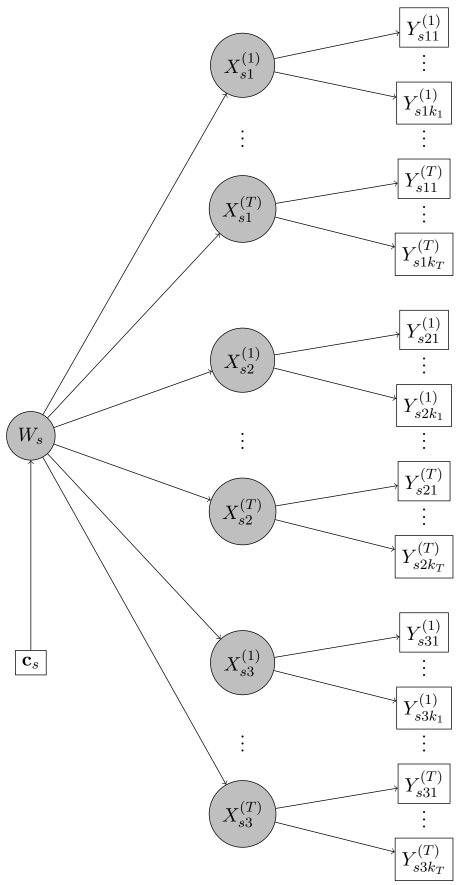
and

$$\begin{aligned}
\widehat{v}_{sdim}^{(t)} &= P\big(X_{sd}^{(t)} = i, W_s = m|\mathbf{Y}_{sd}^{(t)}\big) \\
&= P\big(W_s = m|\mathbf{Y}_s\big) P\big(X_{sd}^{(t)} = i|W_s, \mathbf{Y}_{sd}^{(t)}\big) \\
&= \widehat{u}_{s,m} \frac{P\big(X_{sd}^{(t)} = i|W_s = m\big) P\big(\mathbf{Y}_{sd}^{(t)}|X_{sd}^{(t)} = i\big)}{P\big(\mathbf{Y}_{sd}^{(t)}\big)} \\
&= \widehat{u}_{s,m} \frac{P\big(X_{sd}^{(t)} = i|W_s = m\big) \prod_{h=1}^{k_t} P\big(Y_{sdh}^{(t)}|X_{sd}^{(t)} = i\big)}{\sum_{r=1}^{I} P\big(X_{sd}^{(t)} = r|W_s = m\big) \prod_{h=1}^{k_t} P\big(Y_{sdh}^{(t)}|X_{sd}^{(t)} = r\big)}.
\end{aligned}$$

In the M-step, the expected CDLL is maximized with respect to θ to find the *current updates* for the model parameters - given the expectations computed in the previous step. Standard complete data methods to fit (multinomial) logistic regression models can be used to update the parameter estimates using the *complete data* as if they were observed (Vermunt, 2003).

The E-step and the M-step are iterated until some convergence criterion is fulfil. Subsequently, the hybrid technique implemented in Latent GOLD switches to a Newton-Raphson algorithm. The latter uses analytic first and second order derivatives - see Vermunt and Magidson (2016, Sec. 7.1). The inverse of the negative Hessian, available at convergence, can then be used as an estimate of the observed Information matrix for inference on the model parameters.

Figure 4.1: Path diagram of the proposed multilevel Latent Class model specification.



Note: $X_{s1}^{(t)}$, $X_{s2}^{(t)}$, and $X_{s3}^{(t)}$ correspond to the Level 1 latent variables Knowledge, Application, and Judgment, respectively, at time $t = 1, \ldots, T$. Each of them was measured by $k_t = 10$ indicators per time point $(Y_{sd1}^{(t)}, \ldots, Y_{sdk_t}^{(t)})'$, with $\sum_t k_t = K = 30$. The latent variable W_s at Level 2 defines groups of students based on their likelihood to be in one of the Level 1 latent classes for each Dublin descriptor (K, A, J). Finally, $\mathbf{c}_s$ refers to the demographic and psychological covariates affecting Level 2 class membership probabilities. Light gray circles indicate latent variables, whereas white rectangles indicate observed variables.

4.2 Empirical application

In this section, results from the model fit considering the data presented in Chapter 3 are described. Firstly, the latent structure output (number and characteristics of latent classes) is shown; next, the Level 2 regression output is presented - namely, the effect of covariates on high level class membership probabilities.

Note that in the considered application, $D = 3$ ability dimensions were considered (Knowledge, Application, Judgment), each measured across $T = 3$ time points through a distinct set of $k_t = 10$ multiple-choice Statistics questions, with $\sum_t k_t = 30$. Each question is coded as an ordinal indicator with $q = 3$ categories, indicated with $j = 0, 1, 2$, corresponding to wrong, partially correct, and totally correct answers, respectively.

4.2.1 Latent structure

The definition of the number of latent classes is a ticklish issue in LC analysis in general (Nylund-Gibson et al., 2007; Yang and Yang, 2007). Herein, a set of multilevel latent class models, with a varying number of latent classes (from 1 to 5), both at Level 1 and Level 2, was estimated. These models were then compared using the information criteria based on the number of groups as suggested in Lukočienė and Vermunt (2009). The Sample adjusted Bayesian Information Criterion (SABIC; Sclove, 1987) indicates the model with three latent classes at Level 1 and Level 2 as the best one (see Table 4.1 for details about fit statistics). These results are consistent with those obtained in the application of the second proposal on the same data, which is described in the following chapter. The entropy-based R^2 reveals a high separation between latent classes, i.e. low classification uncertainty, for all the discrete latent variables: $R^2_{ENTR} = 0.94$ for Level 2 group variable, $R^2_{ENTR} = 0.77$ for Knowledge, $R^2_{ENTR} = 0.70$ for Application, and $R^2_{ENTR} = 0.63$ for Judgment latent variables at Level 1.

Table 4.1: Fit statistics for multilevel latent class models.

No. latent classes					
Level 2	Level 1	Log-likelihood	Group-based BIC	Group-based SABIC	Note
1	2	-12640.52	26714.28	25858.87	
	3	-12553.05	27017.08	25876.53	
	4	-12464.23	27317.18	25891.49	
	5	-12385.28	27637.02	25926.20	
2	2	-12441.22	26432.45	25507.33	
	3	-12351.69	26747.07	25527.31	
	4	-12259.28	27055.91	25541.51	
	5	-12170.84	27372.71	25563.66	
3	2	-12409.00	26484.80	25489.98	NC
	3	-12297.60	26771.58	25472.62	
	4	-12227.27	27140.51	25537.40	
	5	-12124.40	27444.38	25537.13	
4	2	-12394.72	26573.02	25508.51	NC
	3	-12274.29	26857.67	25479.50	
	4	-12167.33	27169.28	25477.46	
	5	-12091.07	27542.28	25536.81	

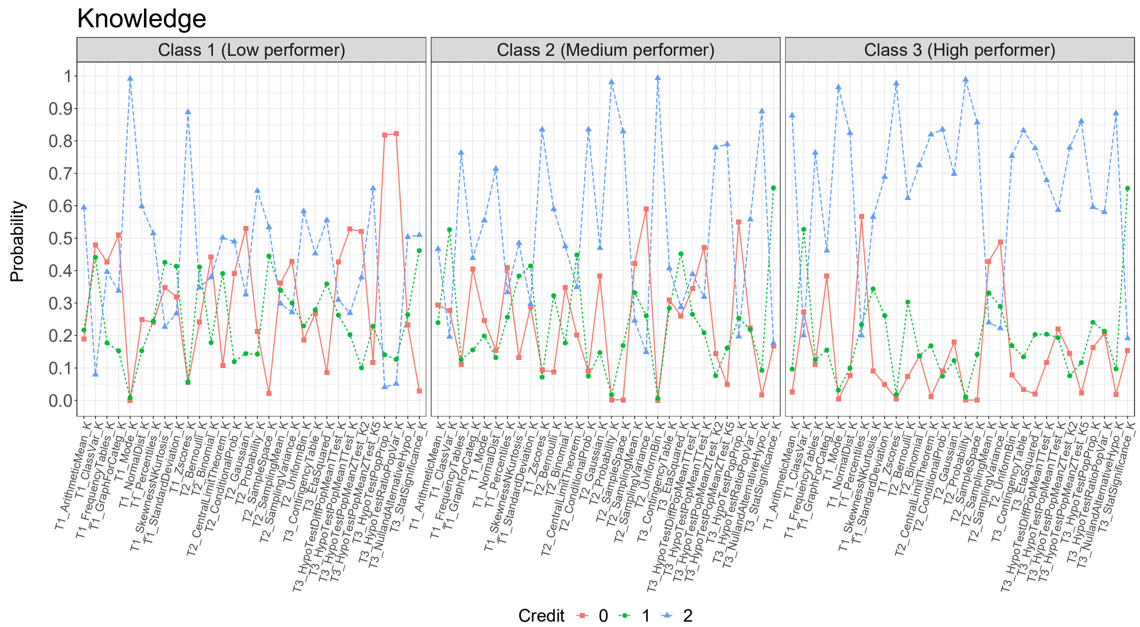
Note: BIC = Bayesian Information Criterion; SABIC = Sample adjusted BIC; NC = Not Converge.

The conditional response probabilities defined in Equation (4.2) can be used to characterize the output latent classes at Level 1. As can be seen in Figures 4.2, 4.3, and 4.4, each one related to one of the considered Dublin descriptors, three ordered latent classes indicate low, medium, and high levels of ability, respectively: the probability of totally correct answers increases moving from Class 1 to Class 3, as well as the number of items with a higher probability of correct answers.

Regarding the weights of latent classes, results show that the largest class for the Knowledge dimension is Class 3 (43%), followed by Class 2 with 38% of the sample. Conversely, for Judgment, Class 1 is the largest class (46%), followed by Class 3 with a class probability of 0.38. Differently, the Application dimension presents almost balanced groups: 39% (Class 1), 31% (Class 2), and 30% (Class 3).

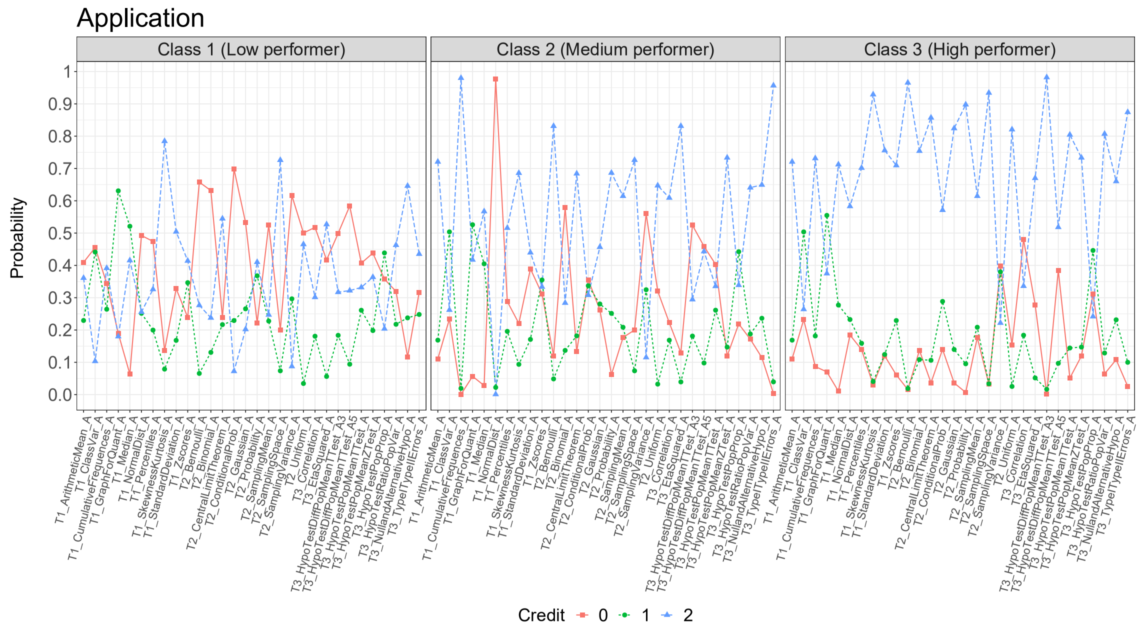
Concerning the Level 2 latent variable, Table 4.2 reports parameter estimates for the low-level class proportions, conditional on higher-level class membership. It can be observed that the discrete latent variable at Level 2 features two well-defined groups of low-performer students (Group 1) and high-performer students (Group 3) for all the Dublin descriptor domains. On the other hand, Group 2 encompasses students with a moderate ability level in Knowledge, and a low or medium level of ability in Application and Judgment. Greater uncertainty in Application and Judgment dimensions was observed for Group 2, with equal class proportion for low (Class 1) and medium (Class 2) levels of ability in Application, and even more distributed proportions between Level 1 latent classes for Judgment.

Figure 4.2: Class profiles for Knowledge according to the response conditional probabilities.



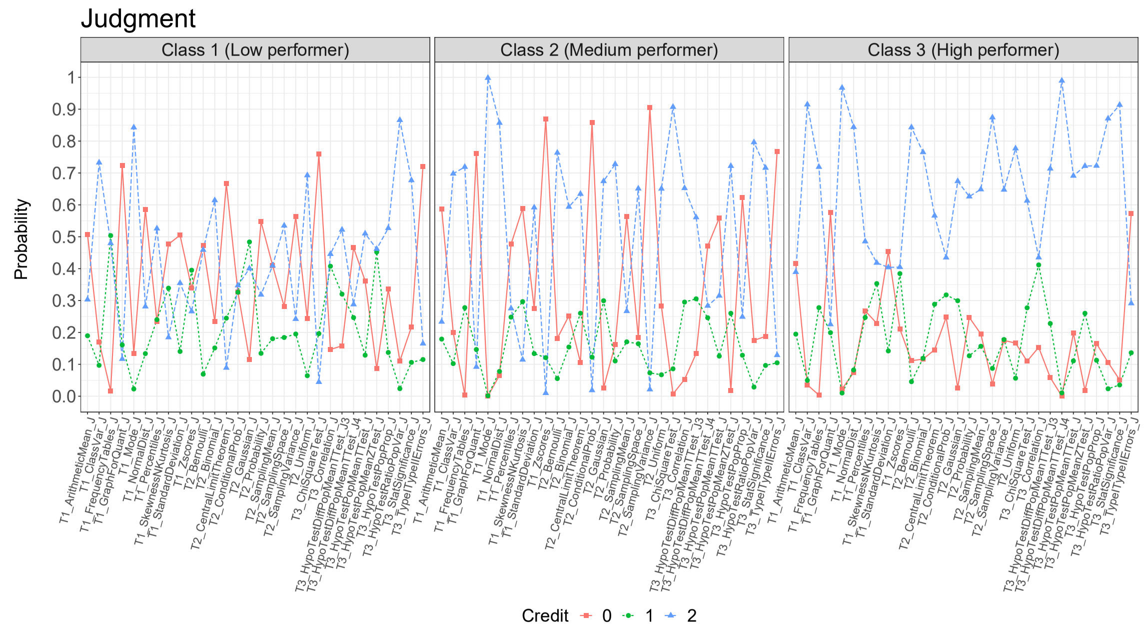
Note: The y-axis indicates the conditional response probability for each category $j = 0, 1, 2$ of the considered indicators whose labels are placed on the x-axis. The three panels differentiate among the emerged latent classes. For each panel, the red line reports the probability of getting a wrong answer ($j = 0$), the green line refers to the probability of getting a partially correct answer ($j = 1$), and the blue line denotes the probability of getting a totally correct answer ($j = 2$).

Figure 4.3: Class profiles for Application according to the response conditional probabilities.



Note: The y-axis indicates the conditional response probability for each category $j = 0, 1, 2$ of the considered indicators whose labels are placed on the x-axis. The three panels differentiate among the emerged latent classes. For each panel, the red line reports the probability of getting a wrong answer ($j = 0$), the green line refers to the probability of getting a partially correct answer ($j = 1$), and the blue line denotes the probability of getting a totally correct answer ($j = 2$).

Figure 4.4: Class profiles for Judgment according to the response conditional probabilities.



Note: The y-axis indicates the conditional response probability for each category $j = 0, 1, 2$ of the considered indicators whose labels are placed on the x-axis. The three panels differentiate among the emerged latent classes. For each panel, the red line reports the probability of getting a wrong answer ($j = 0$), the green line refers to the probability of getting a partially correct answer ($j = 1$), and the blue line denotes the probability of getting a totally correct answer ($j = 2$).

Table 4.2: Low-level class proportion conditional on higher-level class membership averaged over time points. Class weights at both levels are available at the margins.

	Knowledge			Application			Judgment			
Level 2 class	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*	*Weight*
1	0.76	0.22	0.02	0.87	0.12	0.01	0.88	0.10	0.02	*0.25*
2	0.01	0.81	0.18	0.45	0.47	0.08	0.52	0.29	0.19	*0.37*
3	0.01	0.05	0.94	0.01	0.29	0.70	0.11	0.07	0.82	*0.38*
Weight	*0.19*	*0.38*	*0.43*	*0.39*	*0.31*	*0.30*	*0.46*	*0.16*	*0.38*	

4.2.2 Psychological factors and time effects

The statistical significance of the effect of Level 2 covariates on class membership probabilities and of time on Level 1's class proportions, has been evaluated through the Wald test. Note that time effects at Level 1 account for students' ability change according to different Statistics topics, whereas Level 2 covariates provide insights on the impact of demographic and psychological variables on performance.

Results reveal a significant effect of time on Knowledge ($p = 0.04$), and Application ($p = 0.003$), but a non-significant effect on Judgment ($p = 0.17$). More specifically, the output in Table 4.3 shows that the probability of being in Classes 2 and 3 (medium and high performer) rather than in Class 1 (low performer) is lower at time 2 and time 3 in comparison to time 1 (see the negative sign of the corresponding coefficients), pointing out better performance in Knowledge at time 1. On the other hand, students' ability in Application domain increases at time 2 and decreases at time 3, as highlighted from the positive coefficients for Classes 2 and 3 at time 2, and the negative ones at time 3.

All in all, for the Knowledge dimension, results reveal greater inaccuracies and vagueness in the acquisition of theoretical concepts related to probability, bivariate statistics, and hypothesis testing with respect to descriptive statistics, graphs, tables, and normal distribution. In addition, regarding the Application domain, results reveal that students experience more difficulties in bivariate statistics and hypothesis testing with respect to descriptive statistics, whereas they encounter lower difficulty in probability. Conversely, critical ability in Statistics remained stable over time.

Since the topic at each time point varies, the differences in students' performance over time can be explained by the topic complexity and its relevance for the students.

Table 4.3: Parameter estimates for the time-specific intercepts of Level 1 class membership probability's multinomial regressions.

	Time 2: Coefficients		Time 3: Coefficients		Wald statistics	df	p-value
	Class 2	*Class 3*	*Class 2*	*Class 3*			
Knowledge	-5.26	-3.59	-7.55	-1.86	9.90	4	0.04
Application	2.52	0.48	-0.97	-3.92	15.64	4	0.003
Judgment	0.51	1.76	0.43	-0.85	6.36	4	0.17

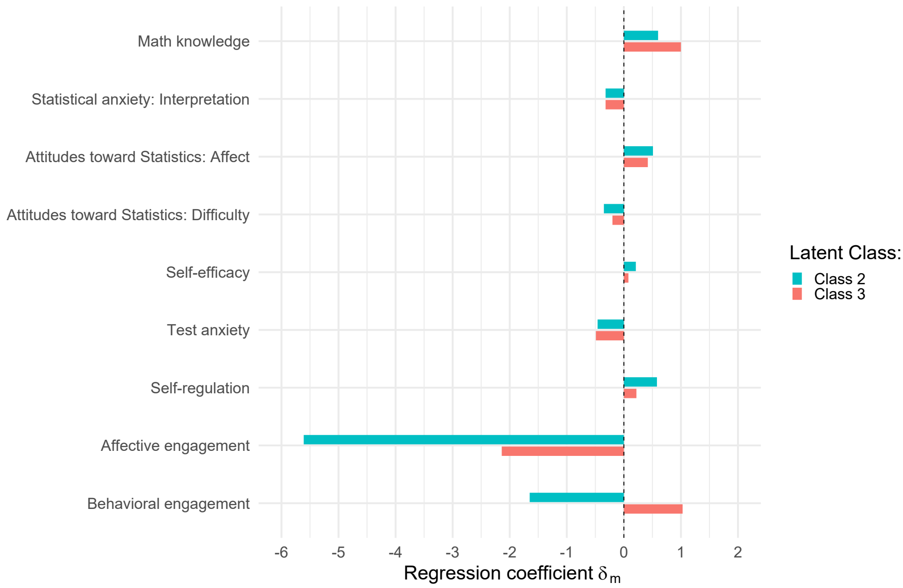
Figure 4.5 displays statistically significant estimates for demographic and psychological covariates. Results highlight that math knowledge, affect attitudes towards Statistics, self-efficacy, and self-regulation positively impact students' performance, increasing their probability of being in Classes 2 and 3 (medium and high performer). Conversely, perceiving Statistics as a difficult subject, test anxiety, and anxiety about interpreting statistical data negatively affect students' performance, as can be seen from the negative sign of the corresponding regression coefficients of Classes 2 and 3.

Students' engagement in Statistics shows a puzzling outcome: the behavioral component (namely learners' observable behaviors related to studying Statistics) has a positive effect on the log-odds of belonging to Class 3 relative to Class 1, but a negative effect on the probability to be in the middle-ability group (Class 2). The affective component of engagement (namely interest, curiosity, and enjoyment in learning statistics) has a negative impact on both the log-odds of belonging to Class 2 or 3 rather than to Class 1. Finally, the following variables have no impact on students' performance: Sex, Academic motivation, Examination anxiety and Fear for asking for help (dimensions of Statistical anxiety), Cognitive competence and Value (dimensions of Attitudes toward Statistics), Cognitive strategies, Academic procrastination, Cognitive engagement.

Looking at the effect of psychological variables on students' performance in Statistics, the obtained results are in line with previous research. Specifically, the positive effect of math knowledge was also described by Chiesi and Primi (2010), and Lavidas et al. (2020). Regarding positive feelings concerning Statistics (i.e., affective attitudes), also other studies (Chiesi and Primi, 2010; Sesé et al., 2015; Sorge, 2001) highlighted their positive impact on students' performance. Moreover, as expected, self-confidence and expectancy for success (i.e., self-efficacy), and the application of useful cognitive and metacognitive strategies to supervise the learning process (i.e., self-regulation) increased the probability of obtaining a good performance. These results also align with the previous literature (Choi, 2005; Zimmerman and Kitsantas, 2014; Zare et al., 2011).

On the other hand, the anxiety occurring when students are required to interpret or make a decision about statistical data, general test anxiety, and considering Statistics as a difficult subject had a detrimental effect

Figure 4.5: Effect of significant psychological covariates on Level 2 class membership probabilities.



on students' performance. The negative impact of these psychological variables was also reported in Tempelaar et al. (2007) and Ghani and Maat (2018), among the others. A surprising result emerges instead for students' engagement in Statistics, underlying the need for further investigations. Indeed, a higher level of behavioral engagement (e.g., study regularly and participate during lessons) is associated with a greater probability of being in the high-performer latent class, but a smaller probability of being in the medium-level ability group. Moreover, the affective component of engagement (namely interest, curiosity, and enjoyment in learning statistics) had an overall negative impact on the performance. This unexpected result may be related to the psychometric scale used to measure students' engagement in Statistics. The latter has not yet been validated in the Italian context, differently from the more classical scales used to assess the other psychological variables.

Chapter 5

The development of a three-step rectangular latent Markov model

This chapter introduces a three-step rectangular latent Markov modeling as an extension of the bias-adjusted three-step latent Markov modeling proposed by Di Mari et al. (2016). In particular, the three-step approach makes it possible to manage different measurement models per time point. The novelty consists in adopting a rectangular formulation of the latent Markov model that enables different numbers of latent classes over time (Anderson et al., 2019). Indeed, as better discussed below, changing over time may lead

Part of the chapter's content is included in the paper:
Fabbricatore, R., Di Mari, R., Bakk, Z., de Rooij, M., & Palumbo, F. (2022). A three-step rectangular latent Markov modeling for advising students in self-learning platforms. In Wiberg, M., Molenaar, D., González, Kim, J.-S., & Hwang, H. (Eds.). *Quantitative Psychology - The 87th Annual Meeting of the Psychometric Society, Bologna, Italy, 2022*. New York: Springer.

to different nature and number of latent classes for which a unique overall definition results too restrictive or redundant. This may be the case when assessing students' abilities, given the complex, multifaceted and transforming nature of learning. Moreover, the rectangular formulation also allows managing another critical issue in education, i.e. students dropout during the learning process. Specifically, in the proposed model, an additional class is considered for time $t = 2, \dots, T$ to explicitly account for dropouts, allowing investigating also the covariate effects on the students' transition to what can be called a "dropout class". As a further innovative element, an item response theory (IRT) parameterization in the measurement part of the model is adopted to take into account item characteristics during the assessment process.

In the following section, the basic latent Markov model is first presented, focusing on the three-step estimation procedure and the extension to rectangular transition matrices. Afterwards, the proposed three-step rectangular latent Markov modeling is outlined. Results from a simulation study provide a preliminary evaluation of the developed bias-adjusted estimator, whereas the real data application illustrates the empirical relevance of the proposed approach for students' ability assessment.

5.1 Latent Markov modeling

As pointed out in Chapter 1, latent Markov models represent the longitudinal extension of models with discrete latent variables, which allow analyzing individuals' transitions across latent states over time. As such, they are particularly helpful to analyze changes in individual or group characteristics over time.

These models assume the presence of a latent process, typically following a first-order Markov chain, which affects the response variables that are repeatedly measured over time. Given the latent process, response variables are considered independent due to the assumption of local independence.

More formally, let $\mathbf{Y}_s^{(t)} = \big(Y_{s1}^{(t)}, \ldots, Y_{sK}^{(t)}\big)'$ be the vector of responses for individual $s = 1, \ldots, N$ on the K indicators measured at time point $t = 1, \ldots, T$, with a realization $\mathbf{y}_s^{(t)}$. Denote with $\mathbf{Y}_s$ the full set of response patterns at all T occasions, and with $X_s^{(t)}$ the categorical latent variable at time t taking value $i = 1, \ldots, I$. Initial and transition probabilities can be respectively formalized through equations

$$P\big(X_s^{(1)} = i\big) = \pi_i^{(1)} \tag{5.1}$$

and

$$P\big(X_s^{(t)} = l | X_s^{(t-1)} = i\big) = \pi_{il}^{(t)}, \tag{5.2}$$

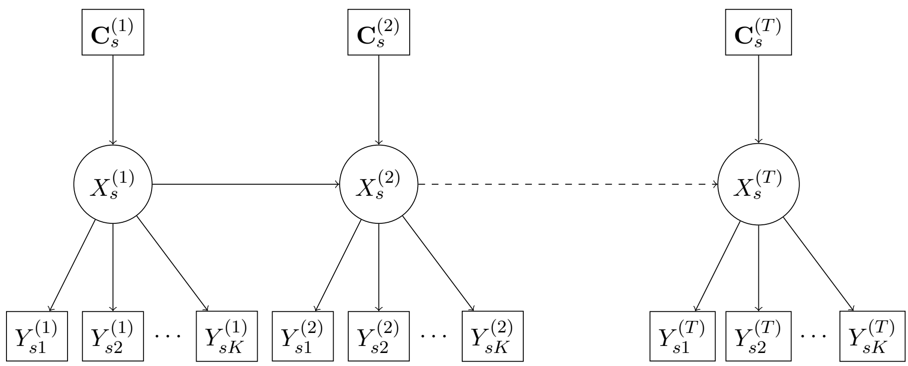
with $l = 1, \ldots, I$, $\sum_{i=1}^{I} \pi_i^{(1)} = 1$ and $\sum_{l=1}^{I} \pi_{il}^{(t)} = 1$. The full set of transition probabilities is collected in a square transition matrix indicated with $\mathbf{\Pi}$. When transition matrices are time-heterogeneous, $T - 1$ different transition matrices $\mathbf{\Pi}^{(2)}, \ldots, \mathbf{\Pi}^{(T)}$ are obtained, with generic element $\pi_{il}^{(t)}$.

Moreover, time-constant and time-varying covariates influencing initial and transition probabilities are allowed. Denote with $\mathbf{C}_s$ the full set of the considered covariates for subject s, being $\mathbf{C}_s^{(t)}$ a sub-vector relative to the specific time t. According to a latent Markov model based on a first-order Markov chain, the probability of having a particular sequence of response configurations at different T time occasions, given the vector of individual covariates, can be expressed as

$$P(\mathbf{Y}_s = \mathbf{y}_s|\mathbf{C}_s) = \sum_{i^{(1)}=1}^{I} \sum_{i^{(2)}=1}^{I} \cdots \sum_{i^{(T)}=1}^{I} P\big(X_s^{(1)} = i^{(1)}|\mathbf{C}_s^{(1)}\big)$$
$$\prod_{t=2}^{T} P\big(X_s^{(t)} = i^{(t)}|X_s^{(t-1)} = i^{(t-1)}, \mathbf{C}_s^{(t)}\big) \prod_{t=1}^{T} P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)}|X_s^{(t)} = i^{(t)}\big). \quad (5.3)$$

The corresponding path diagram is depicted in Figure 5.1.

Figure 5.1: Latent Markov path diagram.



The model comprises two components: the *structural part*, describing the distribution of the latent process, and the *measurement part* that connects the latent state to the response variables at each time point. Thus, looking at the probability defined in Equation (5.3), the structural part refers to the initial state probability $P\big(X_s^{(1)} = i^{(1)}|\mathbf{C}_s^{(1)}\big)$ and the transition probability $P\big(X_s^{(t)} = i^{(t)}|X_s^{(t-1)} = i^{(t-1)}, \mathbf{C}_s^{(t)}\big)$, both conditioned to individual covariates. Conversely, the probability $P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)}|X_s^{(t)} = i^{(t)}\big)$ represents the measurement part of the model, which typically follows a standard latent class model. It can be factorized, because of the local independence assumption, as

$$P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)} | X_s^{(t)} = i^{(t)}\big) = \prod_{k=1}^{K} P\big(Y_{sk}^{(t)} = y_{sk}^{(t)} | X_s^{(t)} = i^{(t)}\big). \tag{5.4}$$

The response variable $Y_{sk}^{(t)}$ can follow a Bernoulli or categorical distribution depending on its binary or ordinal nature, respectively. Accordingly, the conditional-response probabilities $P\big(Y_{sk}^{(t)} = y_{sk}^{(t)} | X_s^{(t)} = i^{(t)}\big)$ can be parameterized by means of the most appropriate link function.

Logistic models can be used to model the effect of the covariates on initial and transition probabilities. In particular, taking the first category as reference, the following parameterization can be used for the initial probability

$$log \frac{P\big(X_s^{(1)} = i | \mathbf{C}_s^{(1)}\big)}{P\big(X_s^{(1)} = 1 | \mathbf{C}_s^{(1)}\big)} = \beta_{0i} + \boldsymbol{\beta}_i' \mathbf{C}_s^{(1)}, \qquad i = 2, \dots, I;$$

where β_{0i} and $\boldsymbol{\beta}_i$ are respectively the intercept and the vector of coefficients for the covariate effects, and

$$log \frac{P\big(X_s^{(t)} = l | X_s^{(t-1)} = i, \mathbf{C}_s^{(t)}\big)}{P\big(X_s^{(t)} = 1 | X_s^{(t-1)} = i, \mathbf{C}_s^{(t)}\big)} = \gamma_{0l} + \gamma_{0il} + \boldsymbol{\gamma}_l' \mathbf{C}_s^{(t)}, \quad i, l = 2, \dots, I$$

for transition probabilities, where γ_{0l} is the intercept for latent state l, γ_{0il} the intercept for the considered transition, and $\boldsymbol{\gamma}_l$ the vector of covariate coefficients. The first class is taken as reference category; accordingly, coefficients related to the first category are set to zero.

Starting from the general formulation of latent Markov models, different constrained versions can be obtained, imposing some restrictions on the measurement and/or the structural part of the model. For example, homogeneous instead of non-homogeneous transition probabilities can be adopted. For a more extensive description of constrained models refer to Bartolucci et al. (2012), and Zucchini and MacDonald (2009). Moreover, a time-variant measurement model can also be considered, as well as a direct effect of covariates on the observed responses (see, for more details, Di Mari and Bakk, 2018).

The estimation of the model parameters is based on maximum likelihood (ML), and it is typically performed in one step by maximizing the following log-likelihood function based on the manifest distribution of the full response pattern $\mathbf{Y}_s$ given the covariates:

$$\ell(\boldsymbol{\eta}) = \sum_{s=1}^{N} \log P\big(\mathbf{Y}_s = \mathbf{y}_s | \mathbf{C}_s\big),$$

where $\boldsymbol{\eta}$ is the vector of free model parameters to be estimated.

Log-likelihood maximization can be performed through the forward-backward algorithm (Baum et al., 1970). It is a special version of the expectation-maximization (EM) algorithm (Dempster et al., 1977), which allows reducing the exponential increase of problem size with the number of time occasions to a linear increase (Zucchini and MacDonald, 2009). Nevertheless, the one-step approach often becomes infeasible when the number of time points, indicators, and covariates, and thus the number of parameters to be estimated, is large (Di Mari et al., 2016). For this reason, step-wise approaches are usually preferred in practice, first addressing the measurement part of the model, and subsequently accounting for the structural part. In this vein, two-step approaches (Bartolucci et al., 2014c) and three-step approaches (Asparouhov and Muthén, 2014; Di Mari et al., 2016) have been proposed. Because the present proposal introduces a three-step procedure, a deeper description of this latter approach is provided in the following subsection.

5.1.1 The three-step approach

The three-step approach breaks down the estimation procedure into the following smaller steps:

Step 1 A simple latent class model without covariates is estimated on the pooled data;

Step 2 State membership is obtained for each individual at all the considered time points;

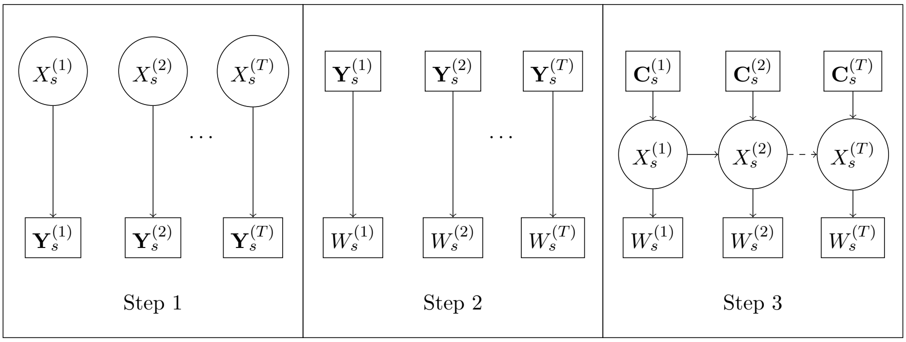
Step 3 Class assignments are used as single indicators in a simple latent Markov model to estimate the structural part of the model (i.e., initial and transition probabilities and covariate effects).

A graphical summary of the three-step approach is provided in Figure 5.2.

It is worth noting that this step-wise approach has the drawback of providing biased estimates in the third step because ignoring the classification error introduced in Step 2. In the context of latent class analysis, two main approaches have been developed to overcome this problem. In particular, Bolck et al. (2004) proposed the BCH correction method based on a weighted logistic regression, whereas Vermunt (2010a) introduced a maximum likelihood-based correction. The latter was generalized to latent Markov models by Di Mari et al. (2016).

As stated above, the second proposal of this thesis consists in extending the bias-adjusted three-step approach to the case of latent Markov models with rectangular transition matrices. Thus, Section 5.2 provides a description of both corrections for the proposed method. Readers interested in a deeper discussion on bias correction methods for latent class analysis and basic latent Markov models can refer to the references cited above.

Figure 5.2: Three-step approach to latent Markov models (adapted from Di Mari et al., 2016).



5.1.2 Rectangular transition matrices

Classical latent Markov models restrict the number of latent states to be equal over time. Nevertheless, this assumption might be too restrictive in some application contexts. To address this issue, Anderson et al. (2019) introduced the rectangular latent Markov models that allow for not time-fixed number of latent states. Accordingly, the categorical latent variable $X_s^{(t)}$ has time-specific I_t support points, producing rectangular transition matrices wherever $I_{t-1} \neq I_t$. In rectangular latent Markov models, Equations (5.1) and (5.2), describing initial and transition probabilities, become

$$P\big(X_s^{(1)} = i | I_1\big) = \pi_{iI_1}^{(1)}$$

and

$$P\big(X_s^{(t)} = l | X_s^{(t-1)} = i, I_{t-1}, I_t\big) = \pi_{ilI_{t-1}I_t}^{(t)},$$

respectively. In addition, the formulation of the measurement part of the model, see Equation (5.4), in the case of a time-varying number of latent classes is

$$P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)} | X_s^{(t)} = i, I_t\big) = \prod_{k=1}^{K} P\big(Y_{sk}^{(t)} = y_{sk}^{(t)} | X_s^{(t)} = i, I_t\big).$$

In latent Markov models with rectangular transition matrices, latent classes can merge, split, or be rearranged over time. As a consequence, when the number of latent states changes, groups with the same label have a different interpretation according to conditional response probabilities. Note that this does not mean that the measurement part of the model is time-varying; indeed, the observed indicators are the same over time.

For parameter estimations, Anderson et al. (2019) propose to consider a penalized likelihood approach due to the very large number of models to be estimated to choose the best configuration for the number of latent states. Indeed, n tested possibilities in classical latent Markov models become n^T when the rectangular formulation is considered. Moreover, they proposed a novel Expectation-Maximization-Markov-Metropolis algorithm to efficiently optimize the penalized likelihood (see, for more details, Anderson et al., 2019). In their proposal, the description of the one-step estimation procedure is driven by the considered application context (classify nations according to well-being), focusing on a latent Markov model without covariates and with normally distributed manifest variables. Very recently, the proposal was extended to a completely general rectangular latent Markov model including covariates and both continuous and categorical outcomes (Russo et al., 2022).

In what follows, a three-step rectangular latent Markov model is introduced, integrating the three-step approach described in Section 5.1.1 with the rectangular formulation of transition matrices. A bias-adjusted approach is proposed for parameter estimation; individual covariates and ordinal observed variables are considered in the empirical application. Regarding the latter, to the best of found knowledge, this work represents the first application of a rectangular latent Markov model to analyze students' assessment data in education.

5.2 A three-step rectangular latent Markov modeling

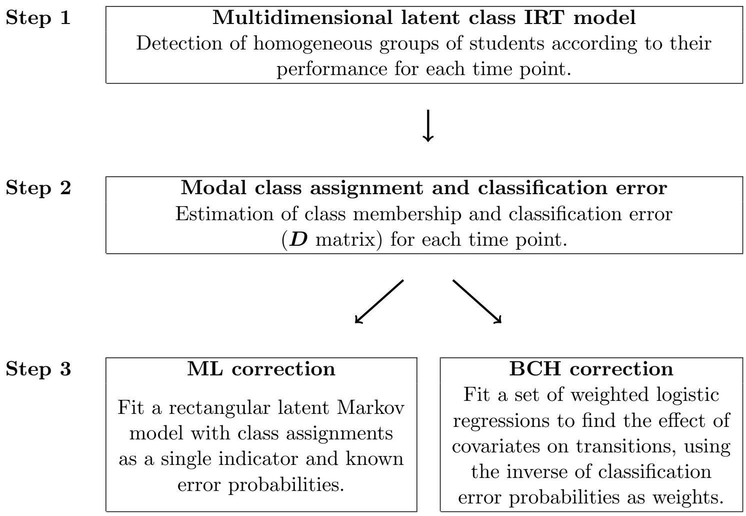
The proposed approach consists of a three-step procedure, running over the following steps: carrying out a multidimensional latent class IRT model at each time point to find homogeneous groups of students according to their ability (Step 1); computing the time-specific class membership and classification error probabilities (Step 2); adopting a bias correction method (BCH or ML-based) to estimate the structural part of the model (Step 3). A brief description of the flow of the proposed three-step approach is provided in Figure 5.3. The remainder of the section details the procedure's steps.

5.2.1 Step 1: Multidimensional latent class IRT model

Step 1 addresses the estimation of the measurement part of the model, which in latent Markov models is typically based on a standard latent class analysis. Herein, an extension that integrates IRT parameterization is considered, leading to the framework of latent class IRT models (Bartolucci, 2007; Bacci et al., 2014). In particular, a between-item multidimensional formulation is exploited to account for the multidimensional nature of the latent trait in the considered context, namely in students' ability assessment.

The multidimensional latent class IRT (MLCIRT) model represents a semi-parametric extension of the traditional IRT model in which both the

Figure 5.3: Flow of the proposed three-step approach.



constraints of unidimensionality and continuous nature of the latent trait are released (Bartolucci, 2007). Given the matrix of students' response patterns, the MLCIRT model allows detecting sub-populations of homogeneous students according to their ability, concurrently accounting for the multidimensional nature of students' ability and item characteristics (e.g., difficulty and discriminating power). Note that the notation used in this section is inspired to the IRT framework; not familiar readers can refer to Bartolucci et al. (2015) for more details.

Formally, the vector $\boldsymbol{\Theta} = \big(\Theta_1, \Theta_2, \dots, \Theta_D\big)'$ of the D latent variables, at each time $t = 1, \dots, T$, follows a discrete distribution with $\boldsymbol{\xi}_1^{(t)}, \boldsymbol{\xi}_2^{(t)}, \dots, \boldsymbol{\xi}_{I_t}^{(t)}$ vectors of support points defining I_t latent classes, where I_t indicates the number of classes at the time t. For any t, $\boldsymbol{\pi}^{(t)} = \pi_1^{(t)}, \dots, \pi_{I_t}^{(t)}$ are the prior probabilities of belonging to latent classes. Specifically, $\pi_i^{(t)} = P\big(\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)$

with $i = 1, \ldots, I_t$, and $\sum_{i=1}^{I_t} \pi_i^{(t)} = 1$. At time $t = 2, ..., T$, the vector of class weights is obtained as $\boldsymbol{\pi}'^{(t)} = \boldsymbol{\pi}'^{(1)} \prod_{h=2}^{t} \boldsymbol{\Pi}^h$, where $\boldsymbol{\Pi}^h$ is the time-specific matrix of transition probabilities of order $I_{h-1} \times I_h$, given the considered working assumptions. It is worth noting that at each time point, the number of latent classes I_t is assumed to be equal for all the latent trait dimensions in order to facilitate the interpretation of the latent classes and improve model parsimony (Bacci et al., 2014). Consequently, individuals belonging to the same latent class share the common profile in terms of all the D latent variables.

The manifest distribution of the individual's response vector $\mathbf{Y}_s^{(t)}$ at time t follows the latent class model specification

$$\begin{aligned} P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)}\big) &= \sum_{i=1}^{I_t} P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)} | \boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big) \pi_i^{(t)} \\ &= \sum_{i=1}^{I_t} \prod_{k=1}^{K_t} P\big(Y_{sk}^{(t)} = y_{sk}^{(t)} | \boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big) \pi_i^{(t)}, \end{aligned} \tag{5.5}$$

where the probability $P\big(Y_{sk}^{(t)} = y_{sk}^{(t)} | \boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)$ is expressed according to an IRT parameterization, and K_t is the number of considered indicators at time t.

Without loss of generality and for the sake of space, the measurement part of the model is presented only referring to the Generalized Partial Credit Model (GPCM; Bacci et al., 2014) among the IRT models. As also pointed out in Chapter 4, a natural way to model the ordered response assuming partial credit for intermediate performance levels, as in the proposed students' assessment, is to use the adjacent-category ordinal logit model or partial credit model (Bartolucci et al., 2015; Eggert and Bögeholz, 2010). Thus, let $\boldsymbol{\theta}^{(t)} = \big(\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_D^{(t)}\big)'$ be a realization of $\boldsymbol{\Theta}$ at time t, and $\theta_{id}^{(t)}$ taking value in $\boldsymbol{\xi}_i^{(t)}$. The response $Y_{sk}^{(t)}$ of the individual $s = 1, \dots, N^{(t)}$ to a generic polytomous item k $\big(k = 1, \dots, K_t\big)$, with J response categories indexed from 0 to $J-1$ and administered at time t, can be parameterized as follows according to the GPCM:

$$
\begin{aligned}
\mathrm{g}\big[P(Y_{sk}^{(t)} = j|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)})\big] &= \log\frac{P(Y_{sk}^{(t)} = j|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)})}{P(Y_{sk}^{(t)} = j-1|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)})} = \\
&= a_k\left(\sum_{d=1}^{D}\delta_{kd}\theta_{id}^{(t)} - b_{kj}\right), \quad j = 1, \dots, J-1;
\end{aligned}
$$

where $\mathrm{g}(\cdot)$ is the local logit link function; δ_{kd} is a dummy variable equal to 1 if the item k measures the latent trait d; a_k and b_{kj} represent the discrimination and the item-step difficulty parameters, respectively. According to a local logit link formulation, the item-step difficulty parameter b_{kj} points out the difficulty of passing from answering category $j-1$ to answering category j to item k. Note that item parameters are constrained to be equal among the latent classes. Moreover, to ensure the identifiability of the model, for each latent trait, it is required that one discriminating index is equal to 1 (usually $a_1 = 1$), and one difficulty parameter is equal to 0 (usually $b_1 = 0$).

Regarding the choice of the number of latent classes I_t, it is a ticklish issue that can be addressed according to some external knowledge or via a model selection process based on a data-driven approach (e.g., comparing models with a different number of classes using the BIC index).

Parameter estimation

Given the number of latent classes I_t, the parameter estimation of the multidimensional latent class IRT model is performed by means of the Maximum Marginal Likelihood (MML) approach (Thissen, 1982).

Let $\boldsymbol{\eta}$ be the vector containing all the free model parameters, the following log-likelihood has to be maximized:

$$\ell(\boldsymbol{\eta}) = \sum_{s=1}^{n} \log\big[P\big(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)}\big)\big] = \sum_{p=1}^{P} n_p \log\big[P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)}\big)\big],$$

where P indicates the total number of distinct response configurations observed in the sample, n_p is the frequency of the p-th response vector $\mathbf{y}_p^{(t)}$, and $P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)}\big)$ defined according to the Equation (5.5). In particular, the model parameters to be estimated are the matrix of ability levels with generic element θ_{id}, the class weights $\pi_i^{(t)}$, and the item discriminant and difficulty parameters.

The maximization of $\ell(\boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$ is obtained using the Expectation-Maximization (EM; Dempster et al., 1977) algorithm that exploits the log-likelihood of the complete data. Given the observed and unobserved data, the complete data log-likelihood CDLL($\boldsymbol{\eta}$) can be formulated as

$$\mathrm{CDLL}(\boldsymbol{\eta}) = \sum_{i=1}^{I_t} \sum_{p=1}^{P} m_{i,p} \log\left\{P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)} | \boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)\pi_i^{(t)}\right\}. \tag{5.6}$$

For estimation purposes, the complete data log-likelihood in Equation (5.6) is decomposed as

$$\mathrm{CDLL}(\boldsymbol{\eta}) = \mathrm{CDLL}(\boldsymbol{\eta}_1) + \mathrm{CDLL}(\boldsymbol{\eta}_2),$$

where $\boldsymbol{\eta}_1$ represent the subvector of $\boldsymbol{\eta}$ containing the free latent class probabilities, and $\boldsymbol{\eta}_2$ all the other free parameters. Specifically,

$$\mathrm{CDLL}(\boldsymbol{\eta}_1) = \sum_{i=1}^{I_t} \sum_{p=1}^{P} m_{i,p} \log\left\{\pi_i^{(t)}\right\} \tag{5.7}$$

and

$$\mathrm{CDLL}(\boldsymbol{\eta}_2) = \sum_{i=1}^{I_t} \sum_{k=1}^{K_t} \boldsymbol{m}_{ik} \log\left\{\boldsymbol{\lambda}_{ik}^{(t)}\right\}, \tag{5.8}$$

where $\boldsymbol{m}_{ik}$ is a column vector with elements $\sum_{p=1}^{P} I\big(Y_{pk}^{(t)} = j\big)m_{i,p}$, with $j = 1,\ldots,J-1$, and $I(\cdot)$ denoting the indicator function equal to 1 if its argument is true (the response of item Y_{pk} is the category j), and 0 otherwise. The vector $\boldsymbol{\lambda}_{ik}^{(t)}$ contains the J conditional probabilities $P\big(Y_{sk}^{(t)} = j|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)$.

The EM algorithm alternates two steps until convergence:

- **E-step.** The missing data $m_{c,p}$ are replaced by conditional expectations to compute the following expected CDLL:

$$\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta})\right] = \sum_{i=1}^{I_t}\sum_{p=1}^{P} \widehat{m}_{i,p} \log\left\{P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)}|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)\pi_i^{(t)}\right\},$$

  where the expected value of $m_{i,p}$ given n_p is computed, for every i and p, as

$$\widehat{m}_{i,p} = n_p \frac{P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)}|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_i^{(t)}\big)\pi_i^{(t)}}{\sum_{h=1}^{I_t} P\big(\mathbf{Y}_p^{(t)} = \mathbf{y}_p^{(t)}|\boldsymbol{\Theta}^{(t)} = \boldsymbol{\xi}_h^{(t)}\big)\pi_h^{(t)}}. \tag{5.9}$$

  Starting from Equation (5.9), the expected frequencies $\sum_{p=1}^{P}\widehat{m}_{i,p}$ and $\widehat{\boldsymbol{m}}_{ik}$ can be determined and substituted in Equation (5.7) and Equation (5.8) to obtain $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_1)\right]$ and $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_2)\right]$, respectively.

- **M-step.** The expected CDLL is maximized with respect to $\boldsymbol{\eta}$ to find the *current updates* for the model parameters - given the expectations computed in the previous step. In particular, the maximization of $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_1)\right]$ have the following explicit solution:

$$\pi_i^{(t)} = \frac{\sum_{p=1}^{P}\widehat{m}_{i,p}}{N^{(t)}}, \qquad i = 2,\ldots,I_t,$$

  whereas the maximization of $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_2)\right]$ is usually implemented through a Fisher-Scoring algorithm (Colombi and Forcina, 2001).

The E- and M-steps are iterated until the difference in the log-likelihoods of two consecutive steps is lower than a threshold (say 10^{-10}). The R package `MultiLCIRT` (Bartolucci et al., 2014a) can be used for the estimation process.

A note about notation: in the following sections, the used notation shifts again to resemble the most typical one in the latent Markov models framework. Therefore, it is important to explicit the connection $P(\boldsymbol{\Theta}_s^{(t)} = \boldsymbol{\xi}_i^{(t)}) = P(X_s^{(t)} = i)$ before moving ahead.

5.2.2 Step 2: Modal class assignment and classification error

Since different measurement models are estimated for each time point, time-specific class membership and classification error probabilities are computed at this step. Once the model parameters are estimated, each observation is assigned to the class corresponding to the highest posterior probability of belonging (modal assignment rule). In particular, the posterior class probability that the individual s belongs to latent class $i = 1, \ldots, I_t$ at time t can be expressed according to the Bayes's theorem as follows:

$$P(X_s^{(t)} = i|\boldsymbol{Y}_s^{(t)} = \boldsymbol{y}_s^{(t)}) = \frac{P(X_s^{(t)} = i)P(\mathbf{Y}_s^{(t)} = \mathbf{y}_s^{(t)}|X_s^{(t)} = i)}{P(\boldsymbol{Y}_s^{(t)} = \boldsymbol{y}_s^{(t)})}.$$

For each time point, modal assignment estimates the predicted class $W_s^{(t)}$ allocating a weight $w_{si}^{(t)} = P(W_s^{(t)} = i|\boldsymbol{Y}_s^{(t)} = \boldsymbol{y}_s^{(t)}) = 1$ if $P(X_s^{(t)} = i|\boldsymbol{Y}_s^{(t)} = \boldsymbol{y}_s^{(t)})$ is the largest posterior probability, and zero weight otherwise.

On the other hand, the conditional probability of the estimated class value conditional on the true one defines the classification error. According to Vermunt (2010a), the overall classification error probabilities are obtained by averaging over all possible observed response configurations at time t, resulting in the time-specific $\boldsymbol{D}^{(t)}$ matrix with elements:

$$P\big(W_s^{(t)} = g | X_s^{(t)} = i\big) = \frac{\frac{1}{N^{(t)}} \sum_{s=1}^{N^{(t)}} P\big(X_s^{(t)} = i | \boldsymbol{Y}_s^{(t)} = \boldsymbol{y}_s^{(t)}\big) w_{sg}^{(t)}}{P\big(X_s^{(t)} = i\big)},$$

where $g, i = 1, \dots, I_t$ and $N^{(t)}$ is the sample size at time t. It is worth noting that the classification error is strongly related to class separation, turning out to be larger in the case of lower separation between classes (Bakk et al., 2013; Vermunt, 2010a).

5.2.3 Step 3 with BCH correction

In Step 3, according to the BCH correction (Bolck et al., 2004), the $\boldsymbol{D}^{(t)}$ matrices computed at the previous step are used in the weighted logistic regressions to find the effect of covariates on initial and transition probabilities. More technically, using a multinomial logistic regression model, the probability of the estimated class membership $W_s^{(t)}$ at time t given the vector $\mathbf{C}_s^{(t)}$ of Q individual covariates can be parameterized as follows:

$$P\big(W_s^{(t)} = g | \boldsymbol{C}_s^{(t)}\big) = \frac{\exp\big(\gamma_{0g}^{(t)} + \sum_{q=1}^{Q} \gamma_{qg}^{(t)} C_{sq}^{(t)}\big)}{\sum_{l=1}^{I_t} \exp\big(\gamma_{0l}^{(t)} + \sum_{q=1}^{Q} \gamma_{ql}^{(t)} C_{sq}^{(t)}\big)}. \tag{5.10}$$

However, the interest is in the relationship between $X_s^{(t)}$ and $\boldsymbol{C}_s^{(t)}$ because simply considering the estimated class membership $W_s^{(t)}$ in Equation (5.10) causes an underestimation of the covariate effects. Thus, in order to model the probability $P\big(X_s^{(t)} = i | \boldsymbol{C}_s\big)$, for $t = 1, \dots, T$, let express the probability $P\big(W_s^{(t)} = g | \boldsymbol{C}_s^{(t)}\big)$ as a linear combination of $P\big(X_s^{(t)} = i | \boldsymbol{C}_s^{(t)}\big)$ considering the classification errors as weights, according to Bolck et al. (2004):

$$P\big(W_s^{(t)} = g | \boldsymbol{C}_s\big) = \sum_{i=1}^{I_t} P\big(X_s^{(t)} = i | \boldsymbol{C}_s^{(t)}\big) P\big(W_s^{(t)} = g | X_s^{(t)} = i\big). \tag{5.11}$$

Let be $e^{(t)}_{sc} = P\big(W^{(t)}_s = g|\boldsymbol{C}_s\big)$, $a^{(t)}_{sg} = P\big(X^{(t)}_s = i|\boldsymbol{C}_s\big)$, and $d^{(t)}_{gc} = P\big(W^{(t)}_s = g|X^{(t)}_s = i\big)$ element of matrices $\boldsymbol{E}^{(t)}$, $\boldsymbol{A}^{(t)}$, and $\boldsymbol{D}^{(t)}$, respectively. The matrix notation of Equation (5.11) is:

$$\boldsymbol{E}^{(t)} = \boldsymbol{A}^{(t)}\boldsymbol{D}^{(t)}.$$

Accordingly, the matrix $\boldsymbol{A}^{(t)}$ with the probabilities of true class membership given individual covariates can be obtained as follows:

$$\boldsymbol{A}^{(t)} = \boldsymbol{E}^{(t)}\boldsymbol{D}^{(t)^{-1}}.$$

Thus, using the entries of the inverse of the $\boldsymbol{D}^{(t)}$ matrix as observation weights during the estimation of the multinomial regression in Equation (5.10), regression parameters referring to the probability $P\big(X^{(t)}_s = i|\boldsymbol{C}_s\big)$ are obtained (Vermunt, 2010a).

Specifically, in the proposed approach, a multinomial regression with individual classification at time 1 as dependent variable allows evaluating the covariate effect on initial probability. On the other hand, to estimate the effect of covariates on all the possible transitions over time, $\sum_{t=1}^{T-1} I_t$ multinomial regressions are required, each one considering the individuals belonging to one of the latent classes emerged at time t as sample (in total I_t sub-samples for each time point) and the corresponding classification at time $t+1$ as dependent variable.

5.2.4 Step 3 with ML correction

Step 3, based on the ML correction, consists of fitting a rectangular latent Markov model with class assignments computed at Step 2 as a single indicator and known error probabilities included in the time-specific $\boldsymbol{D}^{(t)}$ matrices.

The bias-adjusted three-step estimation procedure exploiting the ML correction was introduced in the context of latent Markov models by Di Mari et al. (2016). It extends the work of Vermunt (2010a) on the three-step estimation of LC models. Herein, their proposal is further generalized to the rectangular formulation of latent Markov models.

In the three-step approach, the log-likelihood maximized at Step 3 to estimate the structural part of a latent Markov model is

$$\ell(\boldsymbol{\eta}) = \sum_{s=1}^{N} \log P\big(\mathbf{W}_s|\mathbf{C}_s\big), \tag{5.12}$$

where $\mathbf{W}_s$ is the vector of class membership for all the T occasions, and $\boldsymbol{C}_s$ the full set of the considered covariates for subject s. The probability $P\big(\mathbf{W}_s|\mathbf{C}_s\big)$ can be expressed for rectangular transition matrices as

$$\begin{aligned} P\big(\mathbf{W}_s|\mathbf{C}_s\big) = \sum_{i^{(1)}=1}^{I_1} \sum_{i^{(2)}=1}^{I_2} \cdots \sum_{i^{(T)}=1}^{I_T} & P\big(X_s^{(1)} = i^{(1)}|\mathbf{C}_s^{(1)}\big) \\ & \prod_{t=2}^{T} P\big(X_s^{(t)} = i^{(t)}|X_s^{(t-1)} = i^{(t-1)}, \mathbf{C}_s^{(t)}\big) \\ & \prod_{t=1}^{T} P\big(W_s^{(t)} = g^{(t)}|X_s^{(t)} = i^{(t)}\big), \end{aligned}$$

where $I_1, \ldots, I_T$ are the number of latent states at time $t = 1, \ldots, T$.

Parameters estimation

In standard latent Markov models, the maximization of the third-step log-likelihood defined in Equation (5.12) is typically performed through the Baum-Welch algorithm (Rabiner, 1989), a special case of the EM algorithm which exploits forward and backward probabilities during estimation.

For the proposed rectangular specification, no functions were available for model implementation. Thus, a generalization of the Baum–Welch algorithm is necessary to obtain the parameters' estimate. In what follows, a description of the proposed bias-adjusted estimator for the three-step rectangular latent Markov model is provided, referring to the simplest case of a model without individual covariates. Current developments of the proposed estimation function aim at considering also the covariates' effect on initial and transition probabilities.

Keeping out of consideration the effect of covariates, the third-step log-likelihood expressed in Equation (5.12) reduces to

$$\ell(\boldsymbol{\eta}) = \sum_{s=1}^{N} \log P\big(\mathbf{W}_s\big).$$

Given the observed vector of class membership $\mathbf{W}_s = (W_s^{(1)}, \ldots, W_s^{(T)})$ and the unobserved sequence of latent states $\mathbf{X}_s = \big(X_s^{(1)}, \ldots, X_s^{(T)}\big)$, the complete data log-likelihood $\text{CDLL}\big(\boldsymbol{\eta}\big) = \log\{P\big(\mathbf{W}_s, \boldsymbol{X}_s\big)\}$ can be formulated as

$$\begin{aligned}\text{CDLL}\big(\boldsymbol{\eta}\big) = &\sum_{s=1}^{N}\sum_{i=1}^{I_1} u_{si}^{(1)} \log\Big\{P\big(X_s^{(1)} = i\big)\Big\} + \\ &\sum_{s=1}^{N}\sum_{t=2}^{T}\sum_{i=1}^{I_{t-1}}\sum_{l=1}^{I_t} v_{sil}^{(t)} \log\Big\{P\big(X_s^{(t)} = l | X_s^{(t-1)} = i\big)\Big\} + \\ &\sum_{s=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{I_t} u_{si}^{(t)} \log\Big\{P\big(W_s^{(t)} | X_s^{(t)}\big)\Big\},\end{aligned} \tag{5.13}$$

where $u_{si}^{(t)}$ and $v_{sil}^{(t)}$ can be defined according to the classical EM terminology as follows

$$u_{si}^{(t)} = \begin{cases} 1, & \text{if } X_s^{(t)} = i \\ 0, & \text{otherwise.} \end{cases}, \quad v_{sil}^{(t)} = \begin{cases} 1, & \text{if } X_s^{(t-1)} = i, \quad X_s^{(t)} = l, \\ 0, & \text{otherwise.} \end{cases}$$

As stated before, in order to find the maximum likelihood estimates of the model parameters included in the vector $\boldsymbol{\eta}$, the Baum-Welch algorithm exploits the properties of forward and backward probabilities.

Let the *forward probability* $\alpha_{si}^{(t)}$ be the joint probability $P\big(W_s^{(1)}, \ldots, W_s^{(t)}, X_s^{(t)} = i\big)$, and the *backward probability* $\beta_{si}^{(t)}$ the conditional probability $P\big(W_s^{(t+1)}, \ldots, W_s^{(T)} | X_s^{(t)} = i\big)$. For each time $t = 1, \ldots, T$, it results

$$\alpha_{si}^{(t)} \beta_{si}^{(t)} = P\big(\mathbf{W}_s, X_s^{(t)} = i\big), \qquad i = 1, \ldots, I_t.$$

The vectors $\boldsymbol{\alpha}_s^{(t)}$ and $\boldsymbol{\beta}'^{(t)}_s$ containing the forward and backward probabilities can be expressed, respectively, as

$$\boldsymbol{\alpha}'^{(t)}_s = \boldsymbol{\pi}'^{(1)} \mathbf{P}_s^{(1)} \boldsymbol{\Pi}^{(2)} \mathbf{P}_s^{(2)} \cdots \boldsymbol{\Pi}^{(t)} \mathbf{P}_s^{(t)} = \boldsymbol{\pi}'^{(1)} \mathbf{P}_s^{(1)} \prod_{h=2}^{t} \boldsymbol{\Pi}^{(h)} \mathbf{P}_s^{(h)}$$

$$\boldsymbol{\beta}_s^{(t)} = \boldsymbol{\Pi}^{(t+1)} \mathbf{P}_s^{(t+1)} \boldsymbol{\Pi}^{(t+2)} \mathbf{P}_s^{(t+2)} \cdots \boldsymbol{\Pi}^{(T)} \mathbf{P}_s^{(T)} \mathbf{1} = \Bigg(\prod_{h=t+1}^{T} \boldsymbol{\Pi}^{(h)} \mathbf{P}_s^{(h)} \Bigg) \mathbf{1},$$

where $\mathbf{P}_s^{(t)}$ is the diagonal matrix $I_t \times I_t$ with generic diagonal element equal to the state-dependent probability $P_{si}\big(W_s^{(t)} | X_s^{(t)} = i\big)$. Note that by convention an empty product is the identity matrix, as in the case of $\boldsymbol{\beta}_s^{(T)} = \mathbf{1}$ at $t = T$. For more details about forward and backward probabilities' demonstration refer to Zucchini and MacDonald (2009), among others. From the above-defined relationships follows that

$$\boldsymbol{\alpha}_s'^{(t)}\boldsymbol{\beta}_s^{(t)} = P\big(\mathbf{W}_s\big) = L_T \qquad \text{for } t = 2, \dots, T,$$

providing T different ways to obtain the likelihood L_T, one for each value of t. Nevertheless, the most convenient way to compute L_T is

$$L_T = \boldsymbol{\alpha}_s'^{(T)}\mathbf{1} \tag{5.14}$$

that requires only the computation of forward probabilities, with a general forward recursion defined as

$$\boldsymbol{\alpha}_s'^{(t)} = \boldsymbol{\alpha}_s'^{(t-1)}\boldsymbol{\Pi}^{(t)}\mathbf{P}_s^{(t)}. \tag{5.15}$$

Note that $\mathbf{P}_s^{(t)}$ in Equation (5.15) can be also expressed as a $I_t \times 1$ column vector $\boldsymbol{r}_s^{(t)}$ with generic element $r_{si}^{(t)} = P_{si}\big(W_s^{(t)}|X_s^{(t)} = i\big)$. Let $\boldsymbol{I} = \big(I_1, \dots, I_T\big)$ be the vector with the number of latent states at each time point, $I_{max} = \max\big(\boldsymbol{I}\big)$, and $\boldsymbol{R}_s$ the matrix with dimensions $I_{max} \times T$ containing the T vectors $\boldsymbol{r}_s^{(t)}$.

From a computational point of view, the likelihood in Equation (5.14) requires the product of probabilities, leading to the well-known problem of numerical underflow. To address this issue, a scaling computation of the likelihood is required. In the particular case of the proposed function, the maximum absolute scaling is applied to the vector of forward probabilities at each time t, as described in the Algorithm 1 (see Appendix C). For the sake of clarity, it is worth reminding that $\boldsymbol{\pi}'^{(t)} = \boldsymbol{\pi}'^{(1)} \prod_{h=2}^{t} \boldsymbol{\Pi}^h$.

The forward–backward algorithm for the computation of forward and backward probabilities is described in Algorithm 2 (see Appendix C). Note that $\boldsymbol{H}$ indicates a matrix $N \times T$, with entries $h_s^{(t)}$ equal to 1 if data for subject s are available at time t, and 0 if missing. $\boldsymbol{h}_s'$ is the s-th row vector from $\boldsymbol{H}$. Moreover, define $\boldsymbol{\Gamma}$ as an array $I_{t-1} \times I_t \times (T-1)$

containing the $T-1$ considered transition matrices with generic element $\pi_{il}^{(t)} = P\big(X_s^{(t)} = l | X_s^{(t-1)} = i\big)$.

Forward and backward probabilities are then used in the EM algorithm to maximize the expected CDLL in Equation (5.13) with respect to $\boldsymbol{\eta}$. The complete data log-likelihood is composed of three components related to initial state probabilities, transition probabilities, and state-dependent distributions. Because the state-dependent distributions are considered fixed parameters in the three-step ML approach, only initial and transition probabilities have to be estimated.

In particular, the E-step exploits the following properties of forward and backward probabilities to replace $u_{si}^{(t)}$ and $v_{sil}^{(t)}$ by their conditional expectations given the vector of class membership $\mathbf{W}_s$:

$$\widehat{u}_{si}^{(t)} = P\big(X_s^t = i | \mathbf{W}_s\big) = \alpha_{si}^{(t)} \beta_{si}^{(t)} / L_T$$

$$\widehat{v}_{sil}^{(t)} = P\big(X_s^{t-1} = i, X_s^t = l | \mathbf{W}_s\big) = \alpha_{si}^{(t-1)} \pi_{il}^{(t)} r_{si}^{(t)} \beta_{sl}^{(t)} / L_T.$$

Once the expected value of $u_{si}^{(t)}$ and $v_{sil}^{(t)}$ are obtained, the M-step maximizes the CDLL according to the two sets of parameters of interest, namely initial and transition probabilities. On the other hand, state-dependent distributions are fixed to their estimated values at Step 2.

Thus, the algorithm maximizes:

1) $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_1)\right] = \sum_{s=1}^N \sum_{i=1}^{I_1} \widehat{u}_{si}^{(1)} \log\left\{P\big(X_s^{(1)} = i\big)\right\}$ with respect to $\boldsymbol{\pi}'^{(1)}$ leading to the solution:
$$\pi_i^{(1)} = \widehat{u}_{si}^{(1)} / \textstyle\sum_{i=1}^{I_1} \widehat{u}_{si}^{(1)};$$

2) $\mathbb{E}\left[\text{CDLL}(\boldsymbol{\eta}_2)\right] = \sum_{s=1}^N \sum_{t=2}^T \sum_{i=1}^{I_{t-1}} \sum_{l=1}^{I_t} \widehat{v}_{sil}^{(t)} \log\left\{P\big(X_s^{(t)} = l | X_s^{(t-1)} = i\big)\right\}$

with respect to $\mathbf{\Gamma}$ leading to the solution:

$$\pi_{il} = \sum_{t=2}^{T} \widehat{v}_{sil}^{(t)} / \sum_{t=2}^{T} \sum_{l=1}^{I_t} \widehat{v}_{sil}^{(t)}.$$

The details of the EM algorithm for the proposed method are in the Algorithm 3 (see Appendix C). Given the iterative nature of the EM algorithm, the initialization of the vector of initial probabilities $\boldsymbol{\pi}'^{(t)}$ and transition probability matrices collected in $\mathbf{\Gamma}$ is required. Specifically, in Algorithm 3, initialization is carried out assuming the classification from the second step arises from an observed Markov chain. Note that $\mathbf{W}$ denotes the matrix $N \times T$ with modal class assignment for each subject s at each time point t, and $d\boldsymbol{W}$ the list of the T matrices $N \times I_t$ containing class assignments in a dummy coding. Moreover, let $c\boldsymbol{D}$ be the array $I_{max} \times I_{max} \times T$, including the T time-specific $\boldsymbol{D}^{(t)}$ matrices computed at Step 2 with generic element $d_{gi}^{(t)} = P\big(W_s^{(t)} = g | X_s^{(t)} = i\big)$.

5.3 Simulation study for the developed bias-adjusted estimator

A simulation study was carried out to evaluate the performance of the bias-adjusted estimator for the third step of the proposed approach. Different conditions were considered, mainly concerning class separation and sample size. Note that class separation was herein manipulated by the class–item associations.

The bias in the model parameters estimates was used to compare the estimator's performance under the different conditions. Moreover, CPU time was considered to evaluate how each condition affects the computational time.

5.3.1 Simulation setup

The simulation design is organized as follows:

- **Measurement model.** Three simple latent class models (one per time point) with 3-3-2 latent classes (as emerged in the real data application illustrated in Section 5.4).
- **Number of items.** Ten items for all the considered conditions.
- **Class separation.** Two conditions: Moderate (the response probabilities for the most likely responses were set to 0.8), and Large (the response probabilities for the most likely responses were set to 0.9). The first is indicated with M in the labels, whereas the second is indicated with L.
- **Sample size.** Four conditions: 200, 500, 2000, 10000.
- **Initial probabilities.** Equal size: 1/3, 1/3, 1/3.
- **Transition probabilities.** Persistent Markov chains:

$$\Gamma_1 = \begin{bmatrix} 0.850 & 0.130 & 0.020 \\ 0.100 & 0.800 & 0.100 \\ 0.050 & 0.150 & 0.800 \end{bmatrix};$$

$$\Gamma_2 = \begin{bmatrix} 0.900 & 0.100 \\ 0.600 & 0.400 \\ 0.200 & 0.800 \end{bmatrix}.$$

For each condition, 500 replications were carried out.

5.3.2 Parameters of interest

Parameters of interest include the bias in the estimated parameters, namely initial and transition probabilities, and the CPU time.

Initial probabilities:
The bias for initial probabilities is calculated considering the following relation:

$$\beta_{0i} = log\frac{P\big(X^{(1)} = i\big)}{P\big(X^{(1)} = 1\big)}.$$

Under the given simulation conditions, considering three latent classes at time 1, β_{02} and β_{03} have to be evaluated.

Transition probabilities:
The bias for transition probabilities is calculated considering the following relation:

$$\gamma_{tli} = log\frac{P\big(X^{(t)} = l|X^{(t-1)} = i\big)}{P\big(X^{(t)} = i|X^{(t-1)} = i\big)}.$$

Under the given simulation conditions, six values have to be calculated for the transition matrix from time 1 to time 2 (namely, γ_{121}, γ_{131}, γ_{112}, γ_{132}, γ_{113}, γ_{123}), and three values for the transition matrix from time 2 to time 3 (namely, γ_{221}, γ_{212}, γ_{213}).

CPU Time:
The CPU time required per iteration is stored. Thus, the mean CPU time in seconds was calculated for the full procedure (all the three steps) and for only the third step. The procedure is executed on a 3.00GHz Dell computer with 32GB of RAM running the Windows 11 Pro operating system.

5.3.3 Simulation results

Simulation results support the overall good performance of the proposed third-step bias-adjusted estimator. In particular, Figure 5.4 shows the boxplots of the estimated bias for initial probabilities across the simulated samples under the considered conditions. As can be seen, the variability of the estimated bias distribution becomes smaller as class separation and sample size increase. Thus, large sample size and class separation reduce overestimation problems, enhancing the accuracy of model parameters' estimates. A similar pattern emerges for the estimated bias for transition probabilities, as displayed in Figures 5.5 and 5.6. Small sample size particularly affects parameter estimation, as indicated by the larger standard error of the estimated bias for $n = 200$. However, already with $n = 500$ observations, the accuracy of parameter estimation significantly increases. Moreover, the stronger bias for transition probabilities with respect to initial probabilities in the case of a small sample size ($n = 200$) depends on the presence of very small probabilities in the transition matrix cells that can easily end up in an estimate close to the boundary. Of course, this rarely happens with large samples. Note also that, as stated in Vermunt (2010a), a small separation between classes increases the classification error, affecting the performance of three-step correction methods.

Regarding CPU time, Figure 5.7 shows that it is affected by class separation and sample size, gradually increasing as the class separation gets smaller and sample size increases. Differences in CPU time between the total three-step function and only the third-step reveal that the computational load of the three-step estimation procedure is mainly borne by the third step. The data log-likelihood increases monotonically according to the number of iterations and the algorithm reaches convergence within 20 iterations (see

Figure 5.8 for an example from one of the simulated conditions).

In sum, as expected, the proposed estimator proved to perform well asymptotically, with a larger estimation bias for small samples and lower class separation. These results are in line with those reported in Di Mari et al. (2016) for the bias-adjusted estimator in classical latent Markov models with square transition matrices.

Figure 5.4: Initial probabilities bias.

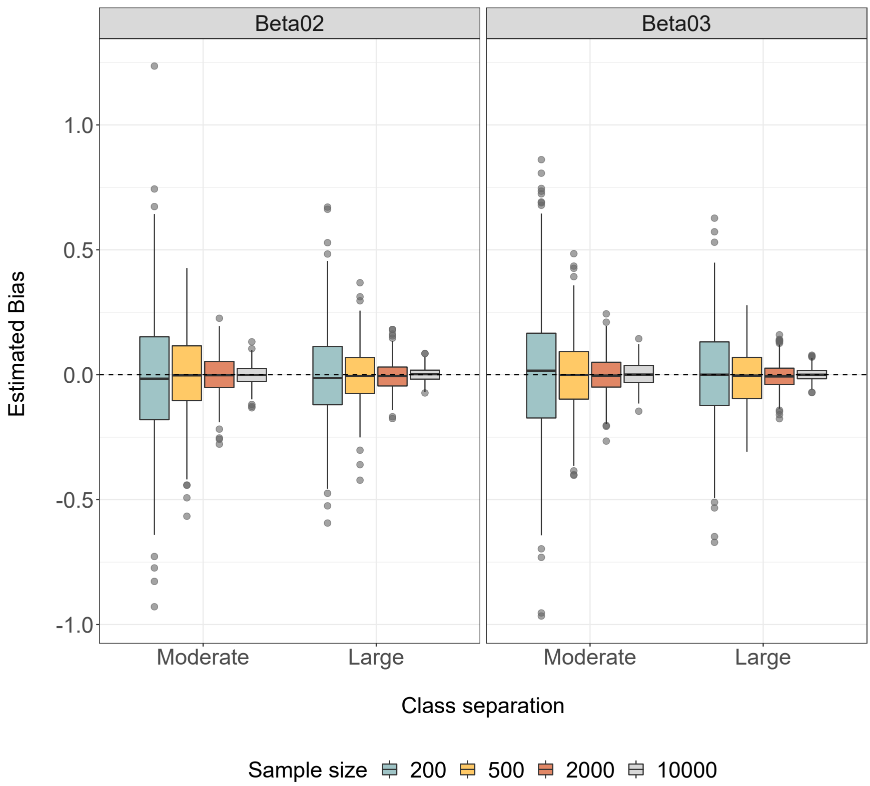


Figure 5.5: Mean and standard error of transition probabilities bias (T1 to T2).

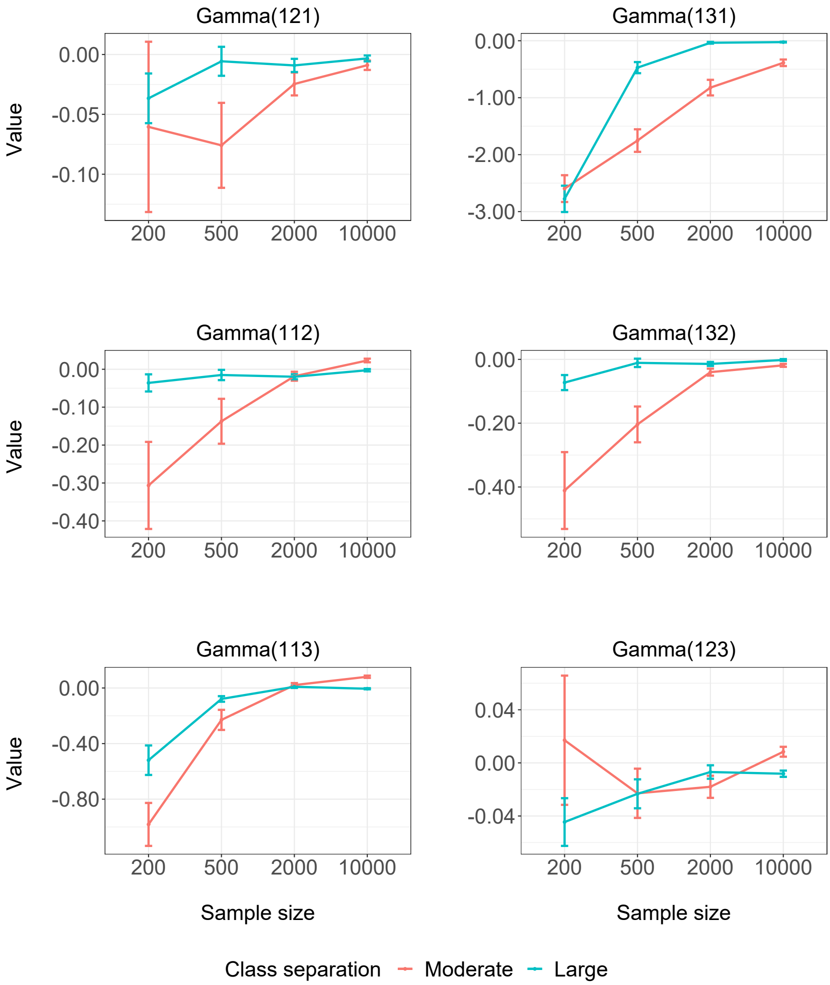


Figure 5.6: Mean and standard error of transition probabilities bias (T2 to T3).

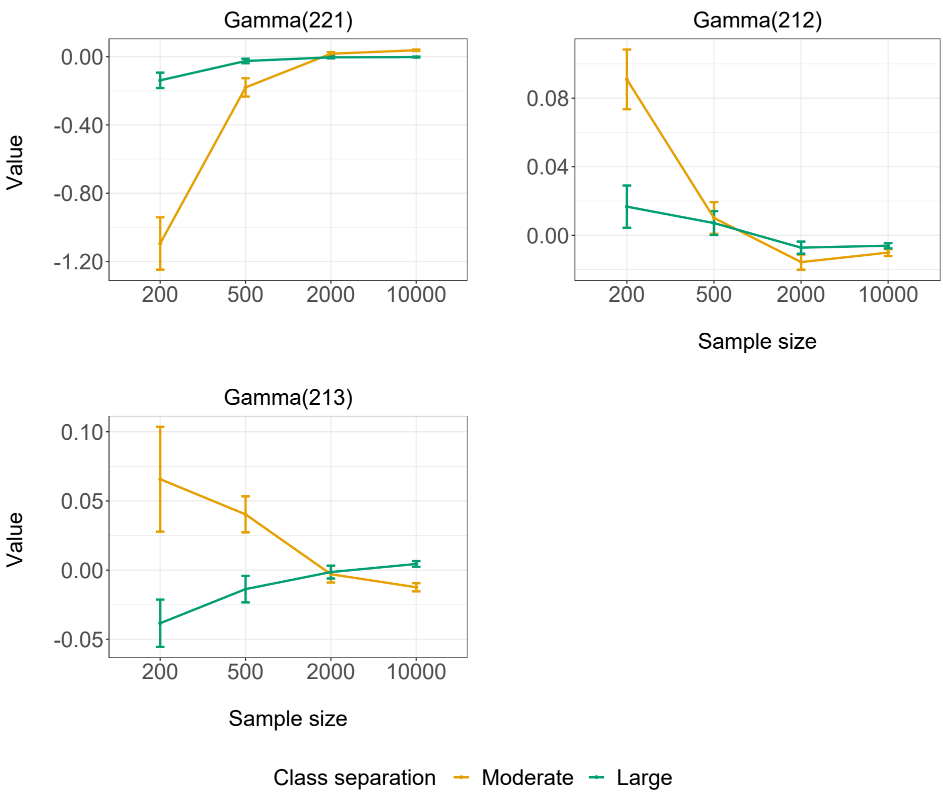
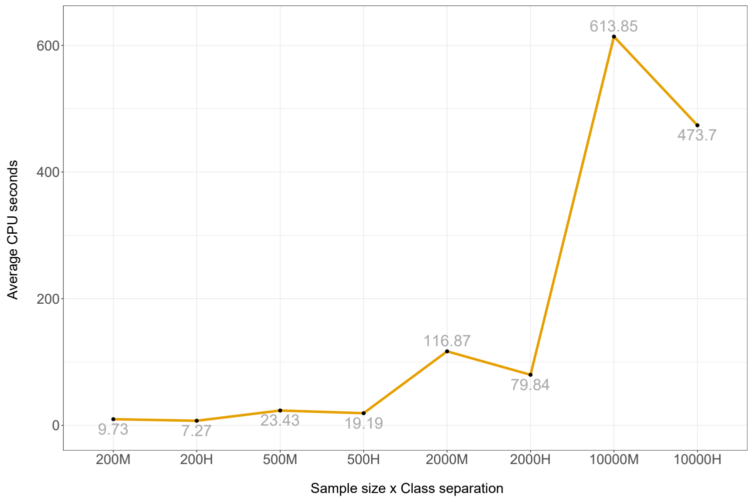


Figure 5.7: CPU time for the total function (a) and the third step (b).

(a)



(b)

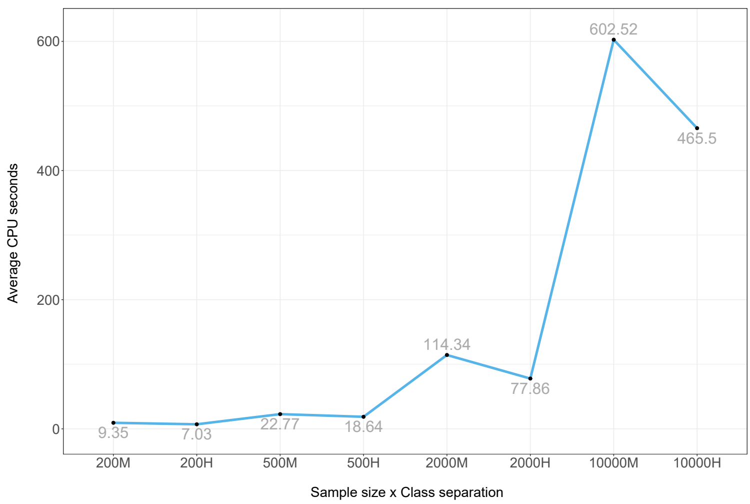
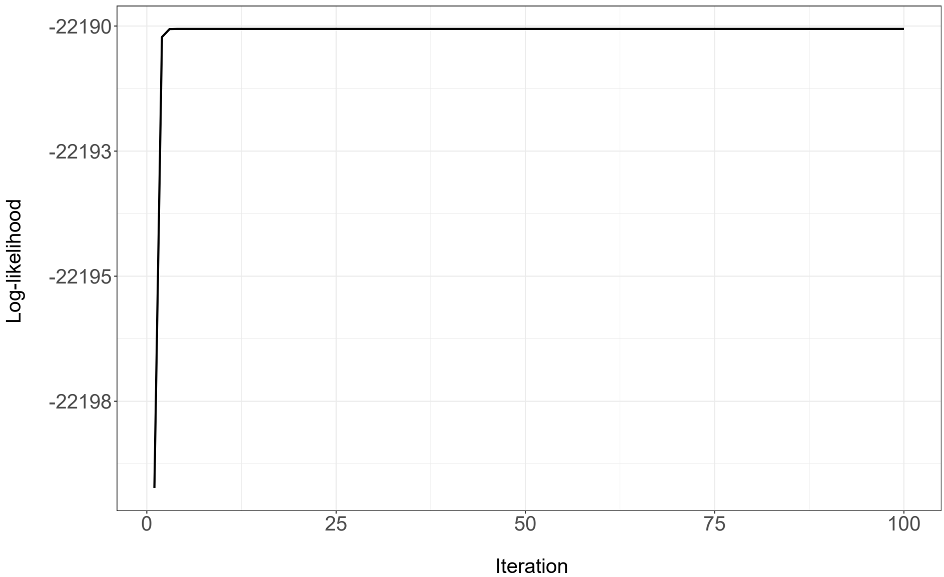


Figure 5.8: Example of log-likelihood series.



5.4 Empirical application

In this section, the three-step rectangular latent Markov approach is used to analyze the complex data structure deriving from the students' assessment proposed in Chapter 3. In particular, this application exploits the BCH correction method for the third step of the estimation procedure. The corresponding R code is provided in the Appendix B.

Firstly, scale comparability across time was ensured in order to compare the classifications obtained at the different time points. It is worth noting that when an IRT model parameterization is exploited to estimate the measurement part of a model, as in the proposed approach, the measurement scale is determined up to an arbitrary linear transformation (Kim and Lee, 2004). Thus, to ensure model identifiability, some constraints have to be imposed, such as standardizing the latent trait distribution or constraining the parameters of a certain item referred to as a "reference item". Consequently, to compare individuals' performance on different test forms, a common metric scale has to be defined by linking the different scales, which is a process called "test equating" (Kolen and Brennan, 2013). When the interest is to track longitudinal trends of students' achievement, one of the most common linking approaches is vertical scaling with a common-item design (Harris, 2007), where there is a set of linking items used across the different test forms. However, as Marengo et al. (2018) pointed out, including common items in a longitudinal evaluation may be unfeasible in some contexts, especially in high-stakes testing, due to a possible learning effect or different test conditions. In such cases, some post-hoc operations can be performed to place scores on various test forms onto a common scale.

To this aim, herein, an IRT factor analysis on the whole set of items was carried out to assess item characteristics and select, for each dimension, the three items (one per time point) with the most similar characteristics in terms of difficulty and discrimination parameters. These "parallel items" can be employed as common items and thus used as reference items for identifiability issues in the multidimensional latent class IRT models at Step 1 to guarantee scale comparability over time. The similarity of the "parallel items" was also tested through a χ^2 test comparing nested IRT factor models where the constrained model has parameters across the reference (parallel) items imposed to be equal. Table 5.1 shows the results for both unidimensional and multidimensional models, all pointing at a not significant difference between the nested models, and thus moving in favor of scale comparability. Given the above results, the parallel items were considered as the reference for model identifiability in the multidimensional latent class IRT models in Step 1.

Table 5.1: IRT factor analysis results: Fit statistics for nested models to test parallel item similarity.

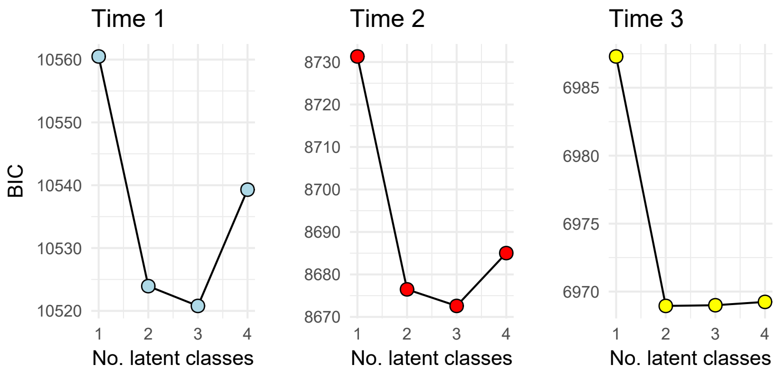
		BIC	χ^2	df	p-value
Knowledge	Constrained	6411.63			
	Unconstrained	6436.25	4.30	6	0.64
Application	Constrained	6891.54			
	Unconstrained	6919.74	0.73	6	0.99
Judgment	Constrained	6735.15			
	Unconstrained	6760.20	3.88	6	0.69
Multidimensional	Constrained	20041.1			
	Unconstrained	20116.4	11.47	18	0.87

Note: Constrained = equal parameters across reference (parallel) items.

The number of latent classes for each time point was defined according to the Bayesian Information Criterion (BIC), pointing at three latent classes at time 1 and time 2, and two latent classes at time 3 (see Figure 5.9).

Looking at class profiles in Figure 5.10, corresponding to the best models in terms of BIC, it can be seen that latent classes are increasingly ordered according to all the latent trait dimensions at each time point. Hence, Class 1, Class 2, and Class 3 indicate low, medium, and high levels of ability, respectively. Moreover, scale comparability allows affirming that students' ability was higher in time 1 (descriptive statistics) than in time 2 and time 3, especially in Knowledge and Judgment. In contrast, a smaller difference in ability levels over time was found for Application. Finally, the level of ability associated to Class 2 at time 3 is very similar to the ability level of

Figure 5.9: BIC values for different number of latent classes at each time point.



Class 2 at time 1. It is worth noting that differences over time in the number and ability characteristics of the latent classes are also influenced by the dropout of some students during learning in addition to the learning topics. As displayed below, the proposed three-step rectangular latent Markov modeling allows for managing these issues effectively, explicitly accounting for class change and dropouts.

Regarding classification error probabilities computed in Step 2, the following $\boldsymbol{D}^{(t)}$ matrices resulted:

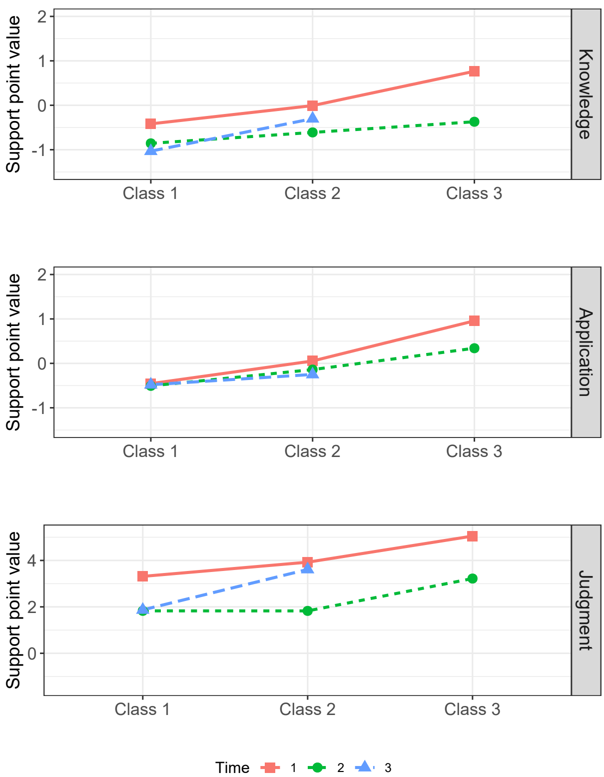
$$\boldsymbol{D}^{(1)} = \begin{bmatrix} 0.834 & 0.165 & 0.000 \\ 0.071 & 0.890 & 0.039 \\ 0.000 & 0.181 & 0.819 \end{bmatrix};$$

$$\boldsymbol{D}^{(2)} = \begin{bmatrix} 0.819 & 0.176 & 0.005 \\ 0.049 & 0.845 & 0.106 \\ 0.002 & 0.049 & 0.950 \end{bmatrix};$$

$$\boldsymbol{D}^{(3)} = \begin{bmatrix} 0.977 & 0.023 \\ 0.031 & 0.969 \end{bmatrix};$$

with the elements on the main diagonals providing evidence for an accurate classification at each time point.

Figure 5.10: Class profiles for the selected models.



Note: Support point value on the y-axis indicates the level of ability of students belonging to the considered latent classes. Accordingly, Class 1, Class 2, and Class 3 indicate low-, medium-, and high-ability learners, respectively.

The inverse of the obtained $\boldsymbol{D}^{(t)}$ matrices was used for the estimation of covariate effects in Step 3. Note that because of the small sample size, a reduced set of the collected covariates was considered in the model.

Results showed that sex, math knowledge, and engagement significantly affect initial classification probabilities, whereas no significant effects were found for statistical anxiety, attitudes toward Statistics, and self-efficacy. In particular, females had a lower probability of being in class 2 ($\gamma_2 = -1.44$, p-value $= 0.058$) and Class 3 ($\gamma_3 = -2.77$, p-value $= 0.001$) with respect to Class 1 than males, highlighting an impairment in the level of ability according to sex at time 1. Moreover, a higher level of math knowledge was associated with a greater probability to be in class 2 ($\gamma_2 = 0.08$, p-value $= 0.03$) and 3 ($\gamma_3 = 0.31$, p-value < 0.001), and thus a medium and high performance in Statistics. Also students' engagement in statistics positively affected students' performance, increasing the probability to be in Class 2 ($\gamma_2 = 0.73$, p-value $= 0.02$) rather than in Class 1.

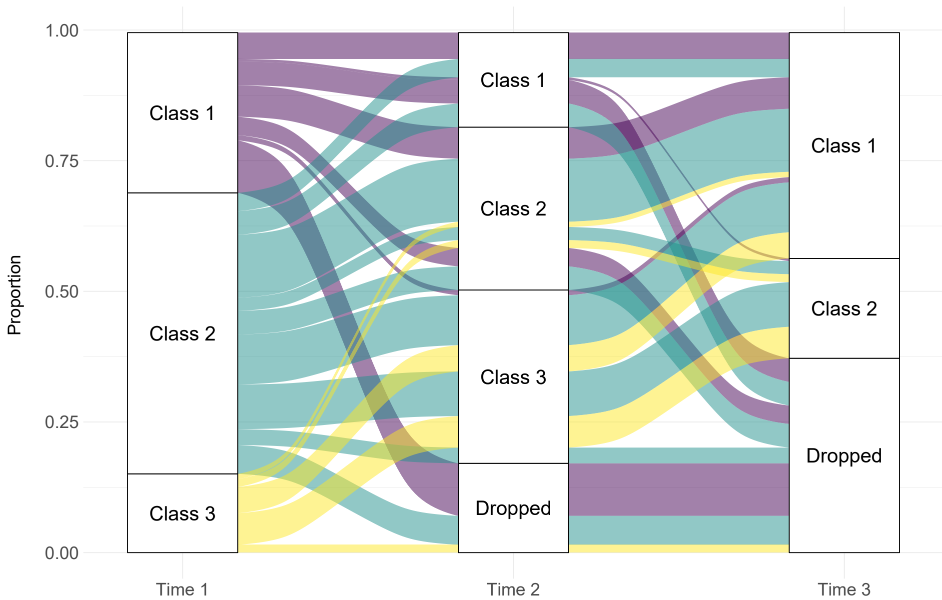
Observed students' transitions over time are depicted in Figure 5.11. Students allocated in Class 1 at time 1 are most at risk of dropout at time 2 and time 3. Students with a high level of ability at time 1 tend to be high performers over time, whereas it is very difficult for a student in Class 1 at time 1 to reach the highest ability level at time 2 and time 3.

The effect of covariates on transition probabilities is reported in Figure 5.12. Note that only the significant effects are reported, using red color for negative effects and green color for positive ones. Moreover, because some students dropped out during learning, an additional class (Dropped) was considered for time 2 and time 3 at this step.

Looking at transitions from time 1 to time 2, it emerged that a lower level of engagement increased the risk of dropout for students belonging to

Class 1, who have low ability levels. Moreover, math knowledge positively affected ability change over time, fostering the transitions of students from Class 2 and Class 3 at time 1 to Class 3 at time 2, namely the class with the highest level of ability.

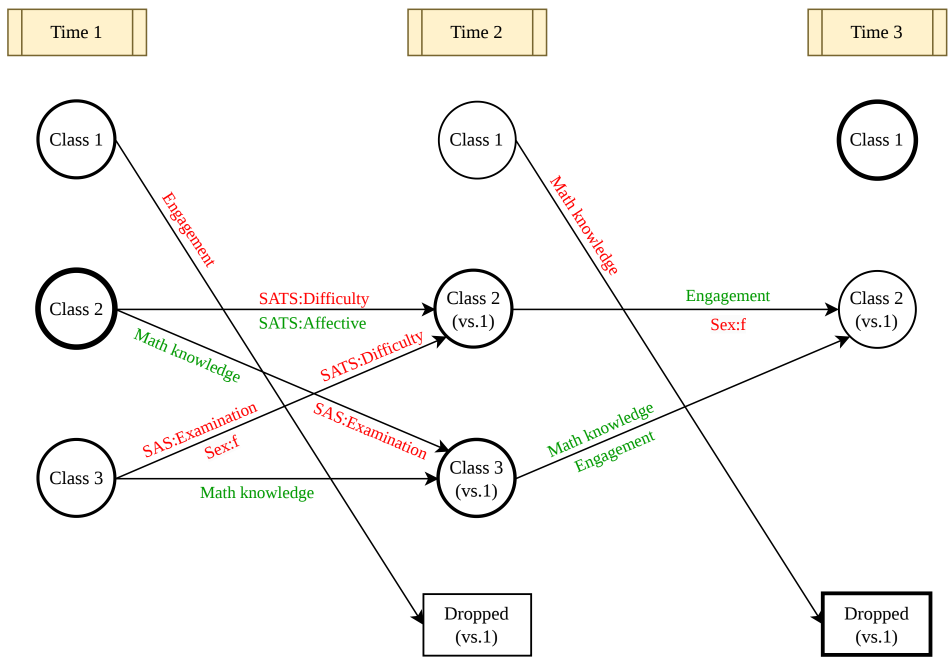
Figure 5.11: Observed students' transitions over time.



Note: The height of the stacked bars at each time is proportional to the class size. The color was assigned according to the classification at time 1. The thickness of the streams connecting the stacked bars between time points is proportional to the number of subjects who have the class identified by both ends of the stream.

In addition, Statistical anxiety and attitudes toward Statistics, although not significantly affecting initial classification probabilities, revealed to have a significant effect on the transitions. Specifically, feeling anxious during a Statistics test and considering Statistics a difficult subject negatively affected students' performance, reducing the probability of moving from Class 2 and Class 3 at time 1 to Class 2 and Class 3 rather than Class 1 at time 2. On the other hand, positive feelings concerning Statistics (affective attitudes) increased the probability of students' transition from Class 2 at time 1 to Class 2 rather than Class 1 at time 2 (positive effect on performance).

Figure 5.12: Significant covariate effects on transitions.



Note: Class 1 always represents the reference class. Negative effects are depicted in red whereas positive effects are in green. Thickness of circles/rectangles represents the proportion of cases most likely assigned to that class. SATS = Attitudes toward Statistics; SAS = Statistical anxiety.

Conversely, transitions from time 2 to time 3 were affected only by math knowledge and engagement among the considered cognitive and psychological variables. This result could be related to the difficulty of the topics at time 3, requiring more basic ability in math and students' effort in studying Statistics to perform well. Regarding the sex variable, results showed that it also affected some transition probabilities, in addition to the initial ones, again underlying impairment in favor of males.

Conclusions

An accurate assessment of students' ability undoubtedly represents a considerable part of a personalized learning activity. However, the derived complex data structure, also produced by the growing use of technology-based learning environments, has brought new challenges for data analysis, especially when the main aim consists in providing students with appropriate feedback on their latent ability level. Indeed, if relying on a comprehensive assessment of students' proficiency lead to the availability of large sets of data on its dimensions and determinants, on the one hand, it requires advanced psychometric modeling to bring the whole latent underlying structure into the light, on the other one.

In this vein, the present contribution proposed two statistical approaches in the framework of non-parametric latent variable models, which allow to detect homogeneous groups of students according to their ability level, concurrently accounting for the effect of individual characteristics on achievements. Specifically, the first proposal exploits a non-standard implementation of multilevel latent class analysis defining a multidimensional latent structure at the low level of the hierarchy to account for the multidimensional nature of students' ability. At the higher level, another discrete latent variable makes it possible to cluster students based on their likelihood to be in one of the lower–level classes for each Dublin descriptor. The obtained two-level clustering turns out helpful to teachers and students to understand points of straight and weaknesses both at a global level and with reference to the learning dimensions. Moreover, the analysis of the time effect on the low-level class membership probabilities provides insights into differences in students' performance according to the considered topics, informing about those that are more complex for students and need to be repeated. Some practical advantages of the proposed strategy are: *(i)* a significantly reduced computational workload compared to parametric approaches, *(ii)* weaker distributional assumptions, *(iii)* the ease of code implementation and *(iv)* the availability of statistical software for parameter estimation.

The second proposal presents a bias-adjusted three-step rectangular latent Markov modeling, either exploiting BCH or ML-based correction methods at the third step. In particular, a new estimation function was built to carry out the ML-based correction, whose good performance was proved by the simulation study. Moreover, the IRT parameterization adopted for the estimation of the measurement part of the model allows to take into account also item characteristics during the assessment process. Thus, the proposed method addresses several typical issues of ability assessment, especially on self-learning platforms, which is based on a different measurement model per time point, different item characteristics (e.g., item difficulty and discrimination), and multiple ability dimensions. Moreover, it allows combining cross-sectional and longitudinal information, identifying students' strengths and weaknesses in comparison with their peers for each topic (cross-sectional) and understanding students' progress over time (longitudinal). In this regard, the rectangular formulation of latent Markov models also accounts for changes leading to different nature and number of latent classes or, as in the presented application, the presence of dropouts. Notably, explicitly accounting for dropouts, and thus for factors affecting students' leaving during learning, provides insights to fine-tune proper actions to reduce that risk. The obtained model parameters can be used to provide students with adaptive feedback during learning at different levels: according to the ability dimensions, the topics, peer performance, and progress over time.

When the topic at each time point varies, as in the described applications, differences in students' performance over time are primarily explained by the considered learning topics. It is important to underline that the treatment of multiple learning topics during the assessment process does not affect the implementation of the proposed statistical models but only requires carefulness for results interpretation. Indeed, individuals' change should not be interpreted as an increase or decrease of the latent trait level over time but rather as a comparison between latent trait levels for the given learning topics. A similar analytics strategy can be found in Nylund-Gibson et al. (2014), where the authors estimated a latent Markov model considering multiple latent class variables that are not repeated measures to study the link between kindergarten readiness profiles and elementary students' reading trajectories.

In both the presented approaches, model parameters can be employed for the prediction of class membership of new individuals according to their response vector and demographic and psychological characteristics. Accordingly, researchers and stakeholders could plan specific interventions to improve learning outcomes, promoting psychological states and skills. It is worth noting that to obtain an accurate prediction of class membership and covariate effects for out-of-sample individuals, reliable estimates of model parameters should be obtained considering representative samples. In this sense, the applications of the proposed approaches, which involve a convenience sample, provide results that are helpful to develop tailored recommendations for students belonging to the considered sample but are not easily generalized to the entire population of psychology students. However, model parameter estimates could also be periodically updated by repeating the same study with other psychology students, thus becoming more accurate as the sample size increases. A further note regarding the application context: herein, the focus on learning statistics in non-STEM degree programs mainly aims to address the Statistical literacy issue that currently represents a relevant topic for society; however, the proposed modeling approaches can also be employed in any different contexts where the latent trait (ability) dimensions and covariates should be defined according to the specific practical or theoretical aim. In this vein, the application in the context of learning Statistics sheds light on the amount of helpful information provided by the models with the aim of encouraging researchers to employ such models when dealing with complex evaluations of students' ability, as the one herein discussed. In addition, future research could also integrate the introduced methods, or some of its advanced versions, in an adaptive learning system where students' clustering is used to develop formative feedback, whereas the analysis of psychological covariate effects drives motivational feedback definition.

References

Adabbo, B., Fabbricatore, R., Iodice D'Enza, A., and Palumbo, F. (2021). Statistics knowledge assessment: an archetypal analysis approach. In Perna, C., Salvati, N., and Schirripa Spagnolo, F., editors, *BoSP SIS2021*, pages 1388–1393. Pearson, Milano.

Adams, R. J., Wilson, M., and Wang, W.-C. (1997a). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1):1–23.

Adams, R. J., Wilson, M., and Wu, M. (1997b). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22(1):47–76.

Akyuz, Y. (2020). Effects of intelligent tutoring systems (ITS) on personalized learning (PL). *Creative Education*, 11(6):953–978.

Albert, J., Cetinkaya-Rundel, M., and Hu, J. (2020). Online statistics teaching and learning. In Howard, J. P. and Beyers, J. F., editors, *Teaching and Learning Mathematics Online*, pages 99–116. Chapman and Hall/CRC, New York.

Alivernini, F. and Lucidi, F. (2008). The Academic Motivation Scale (AMS): Factorial structure, invariance and validity in the Italian context. *Testing, Psychometrics, Methodology in Applied Psychology*, 15(4):211–220.

Allman, E. S., Matias, C., Rhodes, J. A., et al. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37(6A):3099–3132.

Andersen, E. B. (1985). Estimating latent correlations between repeated testings. *Psychometrika*, 50(1):3–16.

Anderson, G., Farcomeni, A., Pittau, M. G., and Zelli, R. (2019). Rectangular latent Markov models for time-specific clustering, with an analysis of the wellbeing of nations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(3):603–621.

Anderson, J. C. and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3):411–423.

Andrade, D. F. and Tavares, H. R. (2005). Item response theory for longitudinal data: population parameter estimation. *Journal of Multivariate Analysis*, 95(1):1–22.

Arminger, G., Stein, P., and Wittenberg, J. (1999). Mixtures of conditional mean-and covariance-structure models. *Psychometrika*, 64(4):475–494.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., and Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4):387–426.

Asparouhov, T. and Muthén, B. (2007). Multilevel mixture models. In Hancock, G. and Samuelsen, K., editors, *Advances in Latent Variable Mixture Models*, pages 27–51. Information Age Publishing, North Carolina.

Asparouhov, T. and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):329–341.

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, 40(4):193–197.

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, 4(1):87–95.

Azevedo, R. and Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, 96:207–210.

Bacci, S., Bartolucci, F., and Gnaldi, M. (2014). A class of Multidimensional Latent Class IRT models for ordinal polytomous item responses. *Communications in Statistics-Theory and Methods*, 43(4):787–800.

Bacci, S., Fabbricatore, R., and Iannario, M. (2022). Multilevel IRT models for the analysis of satisfaction for distance learning during the covid-19 pandemic. *Socio-Economic Planning Sciences*, page 101467.

Baker, R. S. and Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In Veletsianos, G., editor, *Emergence and Innovation in Digital Learning*, pages 83–98. Athabasca University Press, Canada.

Bakk, Z. and Kuha, J. (2021). Relating latent class membership to external variables: An overview. *British Journal of Mathematical and Statistical Psychology*, 74(2):340–362.

Bakk, Z., Tekle, F. B., and Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43(1):272–311.

Bakk, Z. and Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1):20–31.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., and Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440):1375–1386.

Bartholomew, D. J. and Knott, M. (1999). *Latent variable models and factor analysis*, volume 7. Edward Arnold, London.

Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika*, 72(2):141–157.

Bartolucci, F., Bacci, S., and Gnaldi, M. (2014a). MultiLCIRT: An R package for multidimensional latent class item response models. *Computational Statistics & Data Analysis*, 71:971–985.

Bartolucci, F., Bacci, S., and Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*, volume 34. CRC Press, Boca Raton.

Bartolucci, F., Bacci, S., and Mira, A. (2018). On the role of latent variable models in the era of big data. *Statistics & Probability Letters*, 136:165–169.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2012). *Latent Markov models for longitudinal data*. CRC Press, Boca Raton.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2014b). Latent markov models: a review of a general framework for the analysis of longitudinal data with covariates. *Test*, 23(3):433–465.

Bartolucci, F., Montanari, G., and Pandolfi, S. (2014c). A comparison of some estimation methods for latent markov models with covariates. In Gilli, M., Gonzalez-Rodriguez, G., and Nieto-Reyes, A., editors, *Proceedings of COMPSTAT 2014*, pages 531–538.

Barton, M. A. and Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, 1981(1):1–9.

Bauer, D. J. and Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychological Methods*, 9(1):3–29.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.

Ben-Zvi, D. and Makar, K. (2016). International Perspectives on the Teaching and Learning of Statistics. In Ben-Zvi, D. and Makar, K., editors, *The Teaching and Learning of Statistics*, pages 1–10. Springer Cham, Cham.

Bentler, P. M. (1995). *EQS structural equations program manual*, volume 6. Multivariate software Inc, Encino.

Beretvas, S. N. (2011). Cross-classified and multiple-membership models. In Hox, J. and Roberts, J. K., editors, *Handbook of Advanced Multilevel Analysis*, pages 313–334. Routledge, London.

Bergman, L. R., Magnusson, D., and El Khouri, B. M. (2003). *Studying individual development in an interindividual context: A person-oriented approach*. Psychology Press, London.

Bertaccini, B., Grilli, L., and Rampichini, C. (2013). An IRT-MIMIC model for the analysis of university student careers. *Journal of Methodological and Applied Statistics*, 15:95–110.

Bianconcini, S. and Bollen, K. A. (2018). The latent variable-autoregressive latent trajectory model: A general framework for longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5):791–808.

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In Lord, F. M. and Novick, M. R., editors, *Statistical Theories of Mental Test Scores*, pages 397–479. Addison-Wesley, Reading.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1):29–51.

Bolck, A., Croon, M., and Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12(1):3–27.

Bollen, K. A. (1989). *Structural equations with latent variables*, volume 210. John Wiley & Sons, New Jersey.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1):605–634.

Bollen, K. A. and Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions. *Sociological Methods & Research*, 32(3):336–383.

Bollen, K. A. and Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* John Wiley & Sons, New Jersey.

Bollen, K. A. and Zimmer, C. (2010). An Overview of the Autoregressive Latent Trajectory (ALT) Model. In Montfort, K., Oud, J. H., and Satorra, A., editors, *Longitudinal Research with Latent Variables*, pages 153–176. Springer, Berlin.

Bonanomi, A., Olivari, M. G., Mascheroni, E., Gatti, E., and Confalonieri, E. (2018). Using a multidimensional Rash analysis to evaluate the psychometric properties of the motivated strategies for learning questionnaire (MSLQ) among high school students. *Testing, Psychometrics, Methodology in Applied Psychology*, 25(1):83–100.

Bonifay, W. E. (2015). An illustration of the two-tier item factor analysis model. In Reise, S. P. and Revicki, D. A., editors, *Handbook of Item Response Theory Modeling*, pages 207–225. Routledge, London.

Borsboom, D. (2008). Latent variable theory. *Measurement*, 6:25–53.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., and Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. In Brdiczka, O., Chau, P., Carenini, G., Pan, S., and Kristensson, P. O., editors, *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 379–388, New York. ACM.

Bovaird, J. A. and Koziol, N. A. (2012). Measurement models for ordered-categorical indicators. In Hoyle, R. H., editor, *Handbook of structural equation modeling*, pages 495–511. The Guilford Press, New York.

Bray, B. C., Lanza, S. T., and Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1):1–11.

Briggs, D. C. and Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4(1):87–100.

Budé, L., Van De Wiel, M. W., Imbos, T., Candel, M., Broers, N. J., and Berger, M. P. (2007). Students' achievements in a statistics course in relation to motivational aspects and study behaviour. *Statistics Education Research Journal*, 6(1):5–21.

Buhl, M. and Andreasen, L. B. (2018). Learning potentials and educational challenges of massive open online courses (MOOCs) in lifelong learning. *International Review of Education*, 64:151–160.

Cagnone, S., Moustaki, I., and Vasdekis, V. (2009). Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology*, 62(2):401–415.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75(4):581–612.

Cappé, O., Moulines, E., and Rydén, T. (2009). *Inference in Hidden Markov Models*. Springer, New York.

Carannante, M., Davino, C., and Vistocco, D. (2019). Learning and engagement assessment in MOOCs using multivariate methods and models. In Calise, M., Delgado Kloos, C., Mongenet, C., Reich, J., Ruipérez-Valiente, J. A., Shimshon, G., Staubitz, T., and Wirsing, M., editors, *EMOOCs-WIP*, pages 27–32.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.

Chiesi, F. and Primi, C. (2009). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the Survey of Attitudes toward Statistics (SATS). *Learning and Individual Differences*, 19(2):309–313.

Chiesi, F. and Primi, C. (2010). Cognitive and non-cognitive factors related to students'statistics achievement. *Statistics Education Research Journal*, 9(1):6–26.

Chiesi, F., Primi, C., and Carmona, J. (2011). Measuring statistics anxiety: Cross-country validity of the Statistical Anxiety Scale (SAS). *Journal of Psychoeducational Assessment*, 29(6):559–569.

Choi, N. (2005). Self-efficacy and self-concept as predictors of college students' academic performance. *Psychology in the Schools*, 42(2):197–205.

Christenson, S., Reschly, A. L., Wylie, C., et al. (2012). *Handbook of Research on Student Engagement*, volume 840. Springer, New York.

Chung, H., Park, Y., and Lanza, S. T. (2005). Latent transition analysis with covariates: Pubertal timing and substance use behaviours in adolescent females. *Statistics in Medicine*, 24(18):2895–2910.

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In Buckingham Shum, S., Gasevic, D., and Ferguson, R., editors, *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pages 134–138, New York. ACM.

Cokley, K. O., Bernard, N., Cunningham, D., and Motoike, J. (2001). A psychometric investigation of the academic motivation scale using a United States sample. *Measurement and Evaluation in Counseling and Development*, 34(2):109–119.

Collins, L. M. and Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*, volume 718. John Wiley & Sons, Hoboken.

Collins, L. M. and Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27(1):131–157.

Colombi, R. and Forcina, A. (2001). Marginal regression models for the analysis of positive association of ordinal response variables. *Biometrika*, 88(4):1007–1019.

Conole, G., Gašević, D., Long, P., and Siemens, G. (2011). *Proceedings of the 1st international conference on learning analytics and knowledge.* ACM, New York.

Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* ERIC, Washington.

Croon, M. A. and Bolck, A. (1997). *On the use of factor scores in structural equations models.* WORC Paper. Oldendorff Research Institute, Tilburg.

Cutler, A. and Breiman, L. (1994). Archetypal analysis. *Technometrics*, 36(4):338–347.

Dayton, C. M. and Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83(401):173–178.

de Barba, P. G., Kennedy, G. E., and Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*, 32(3):218–231.

De Boeck, P. and Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*, volume 10. Springer, New York.

de Chiusole, D., Stefanutti, L., Anselmi, P., and Robusto, E. (2020). Stat-knowlab. assessment and learning of statistics with competence-based knowledge space theory. *International Journal of Artificial Intelligence in Education*, 30(4):668–700.

De La Torre, J. and Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3):333–353.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Desmarais, M. C. and Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1):9–38.

Di Mari, R. and Bakk, Z. (2018). Mostly harmless direct effects: A comparison of different latent markov modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3):467–483.

Di Mari, R., Bakk, Z., Oser, J., and Kuha, J. (2022). A two–step estimator for multilevel latent class analysis with covariates. *Submitted to journal.*

Di Mari, R., Bakk, Z., and Punzo, A. (2020). A random-covariate approach for distal outcome prediction with latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(3):351–368.

Di Mari, R., Oberski, D. L., and Vermunt, J. K. (2016). Bias-adjusted three-step latent Markov modeling with covariates. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5):649–660.

Doignon, J. P. and Falmagne, J. C. (1999). A practical application: The ALEKS system. In Falmagne, J. C., Albert, D., Doble, C., Eppstein, D., and Hu, X., editors, *Knowledge Spaces*, pages 9–10. Berlin.

Doignon, J. P. and Falmagne, J. C. (2016). Knowledge spaces and learning spaces. In Batchelder, W., Colonius, H., and Dzhafarov, E., editors, *New Handbook of Mathematical Psychology*, pages 274–321. Cambridge University Press, Cambridge.

Dolan, C. V. and van der Maas, H. L. (1998). Fitting multivariage normal finite mixtures subject to structural equation modeling. *Psychometrika*, 63(3):227–253.

Dziak, J. J., Bray, B. C., Zhang, J., Zhang, M., and Lanza, S. T. (2016). Comparing the performance of improved classify-analyze approaches for distal outcomes in latent profile analysis. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 12(4):107–116.

D'Mello, S. K. (2021). Improving student engagement in and with digital learning technologies. In Vincent-Lancrin, S., editor, *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, pages 79–104. OECD Publishing, Paris.

D'Mello, S. K. and Graesser, A. C. (2014). Feeling, thinking, and computing with affect-aware learning. In Calvo, R., D'Mello, S., Gratch, J., and Kappas, A., editors, *The Oxford Handbook of Affective Computing*, pages 419–434. Oxford University Press, Oxford.

Edwards, M. C., Wirth, R. J., Houts, C. R., and Xi, N. (2012). Categorical data in the structural equation modeling framework. In Hoyle, R. H., editor, *Handbook of Structural Equation Modeling*, pages 195–208. The Guilford Press, New York.

Eggert, S. and Bögeholz, S. (2010). Students' use of decision-making strategies with regard to socioscientific issues: An application of the Rasch partial credit model. *Science Education*, 94(2):230–258.

Eid, M. (1996). Longitudinal confirmatory factor analysis for polytomous item responses: Model definition and model selection on the basis of stochastic measurement theory. *Methods of Psychological Research Online*, 1(4):65–85.

Elsom-Cook, M. (1993). Student modelling in intelligent tutoring systems. *Artificial Intelligence Review*, 7(3-4):227–240.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56(3):495–515.

Eysenck, M. W., Derakshan, N., Santos, R., and Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, 7(2):336–353.

Fabbricatore, R., Parola, A., Pepicelli, G., and Palumbo, F. (2021). A latent class approach for advising in learning statistics: Implementation in the ALEAS system. In Limone, P. and Di Fuccio, R., editors, *Proceedings of the First Workshop on Technology Enhanced Learning Environments for Blended Education*, volume 2817. CEUR Workshop Proceedings.

Fabbricatore, R., Parola, A., Pepicelli, G., and Palumbo, F. (2022a). Academic procrastination scale: psychometric evidence for the Italian version based on CFA and IRT models. *Under review.*

Fabbricatore, R., Parola, A., Pepicelli, G., and Palumbo, F. (2022b). Understanding undergraduate statistics performance through a PLS-PM approach: the role of math knowledge, test anxiety and attitudes toward Statistics. *Statistica Applicata-Italian Journal of Applied Statistics*, (Accepted).

Fadeev, A. et al. (2019). Redesigning contemporary learning in literature, history and culture through online learning environments. In Roceanu, I., editor, *Conference proceedings of ≫eLearning and Software for Education≪ (eLSE)*, volume 15, pages 223–228. ‘Carol I’ National Defence University Publishing House, Bucharest.

Fadel, C., Holmes, W., and Bialik, M. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* The Center for Curriculum Redesign, Boston.

Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., and Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students’ mathematical strategies. *Applied Measurement in Education*, 29(2):144–159.

Feingold, A., Tiberio, S. S., and Capaldi, D. M. (2014). New approaches for examining associations with latent categorical variables: Applications to substance abuse and aggression. *Psychology of Addictive Behaviors*, 28(1):257–267.

Ferguson, G. A. (1942). Item selection by the constant process. *Psychometrika*, 7(1):19–29.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317.

Fielding, A. and Goldstein, H. (2006). *Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review.* DfES, Nottingham.

Finch, W. H. and French, B. F. (2014). Multilevel latent class analysis: Parametric and nonparametric models. *The Journal of Experimental Education*, 82(3):307–333.

Fischer, G. H. (2001). Gain scores revisited under an IRT perspective. In Boomsma, A., Duijn, A. J. M., and Duijn, Snijders, T. A. B., editors, *Essays on Item Response Theory*, pages 43–68. Springer, New York.

Fox, J. P. (2001). *Multilevel IRT: A Bayesian Perspective on Estimating Parameters and Testing Statistical Hypotheses.* Ph.D. thesis, University of Twente, Enschede.

Fox, J. P. and Glas, C. A. (2001). Bayesian estimation of a multilevel irt model using gibbs sampling. *Psychometrika*, 66(2):271–288.

Gadanidis, G. (2017). Artificial intelligence, computational thinking, and mathematics education. *International Journal of Information and Learning Technology*, 34(2):133–139.

Galli, S., Chiesi, F., and Primi, C. (2008). The construction of a scale to measure mathematical ability in psychology students: An application of the Rasch Model. *Testing, Psychometrics, Methodology in Applied Psychology*, 15(1):3–18.

Gamage, S. H., Ayres, J. R., and Behrend, M. B. (2022). A systematic review on trends in using Moodle for teaching and learning. *International Journal of STEM Education*, 9(9):1–24.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In Aigner, D. and Goldberger, A., editors, *Latent Variables in Socio-Economic Models*. North-Holland, Amsterdam.

Ghani, F. H. A. and Maat, S. M. (2018). Anxiety and achievement in statistics: A systematic review on quantitative studies. *Creative Education*, 9(14):2280–2290.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., and Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31(1):4–19.

Gibbons, R. D. and Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57(3):423–436.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 24(3):229–252.

Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology*, 33(2):234–246.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part IA modified latent structure approach. *American Journal of Sociology*, 79(5):1179–1259.

Gorsuch, R. L. (1983). *Factor Analysis*. Lawrence Erlbaum Associates, Hillsdale.

Gorter, R., Fox, J., Riet, G. T., Heymans, M., and Twisk, J. (2020). Latent growth modeling of IRT versus CTT measured longitudinal latent variables. *Statistical Methods in Medical Research*, 29(4):962–986.

Graf, S., Ives, C., Rahman, N., and Ferri, A. (2011). AAT: a tool for accessing and analysing students' behaviour data in learning systems. In Conole, G., Gašević, D., Long, P., and Siemens, G., editors, *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pages 174–179, New York. ACM.

Green Jr, B. F. (1952). Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association*, 47(257):71–76.

Grün, B. and Zeileis, A. (2009). Automatic generation of exams in R. *Journal of Statistical Software*, 29(10):1–14.

Gudeva, L. K., Dimova, V., Daskalovska, N., and Trajkova, F. (2012). Designing descriptors of learning outcomes for higher education qualification. *Procedia-Social and Behavioral Sciences*, 46:1306–1311.

Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: local dependence models. *Sociological Methods & Research*, 16(3):379–405.

Hagenaars, J. A. and McCutcheon, A. L. (2002). *Applied latent class analysis.* Cambridge University Press, Cambridge.

Hanna, D. and Dempster, M. (2009). The effect of statistics anxiety on students' predicted and actual test scores. *The Irish Journal of Psychology*, 30:201–209.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., and Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, 48:615–625.

Harley, J. M., Lajoie, S. P., Frasson, C., and Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education*, 27(2):268–297.

Harris, D. J. (2007). Practical issues in vertical scaling. In Dorans, N. J., Pommerich, M., and Holland, P. W., editors, *Linking and aligning scores and scales*, pages 233–251. Springer, New York.

Hatziapostolou, T. and Paraskakis, I. (2010). Enhancing the impact of formative feedback on student learning through an online feedback system. *Electronic Journal of E-learning*, 8(2):111–122.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Sage Publications Inc, New York.

Heller, J., Stefanutti, L., Anselmi, P., and Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80(4):995–1019.

Henry, K. L. and Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2):193–215.

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., and Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, 66:4–15.

Hill, C. D. (2006). *Two models for longitudinal item response data.* PhD thesis, The University of North Carolina at Chapel Hill, Chapel Hill.

Hoijtink, H. (1995). Linear and repeated measures models for the person parameters. In Fischer, G. H. and Molenaar, I. W., editors, *Rasch models*, pages 203–214. Springer, New York.

Holmes, W., Anastopoulou, S., Schaumburg, H., and Mavrikis, M. (2018). *Technology-enhanced personalised learning: Untangling the evidence*. Robert Bosch Stiftung GmbH, Stuttgart.

Holmes, W., Bialik, M., and Fadel, C. (2019). *Artificial intelligence in education*. Center for Curriculum Redesign, Boston.

Holzinger, K. J. and Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1):41–54.

Honicke, T. and Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, 17:63–84.

Hsu, M. H., Chen, P. S., and Yu, C. S. (2021). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, pages 1–12.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *The Journal of Experimental Education*, 84(1):175–196.

Huang, F. L. and Cornell, D. G. (2016). Using multilevel factor analysis with clustered data: Investigating the factor structure of the positive values scale. *Journal of Psychoeducational Assessment*, 34(1):3–14.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., and Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior*, 18(6):667–683.

Istat (2022). Statistica! dire, fare, capire. https://www.istat.it/storage/Dire-Fare-Capire/Cap1.pdf.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., and Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1):6–47.

Jedidi, K., Jagpal, H. S., and DeSarbo, W. S. (1997). STEMM: A general finite mixture structural equation model. *Journal of Classification*, 14(1):23–50.

Jeon, M. and Rabe-Hesketh, S. (2016). An autoregressive growth model for longitudinal item analysis. *Psychometrika*, 81(3):830–850.

Jeon, S., Lee, J., Anthony, J. C., and Chung, H. (2017). Latent class analysis for multiple discrete latent variables: A study on the association between violent behavior and drug-using behaviors. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(6):911–925.

Johnson, M. and Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education*, 14(2):1–15.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2):183–202.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2):109–133.

Jöreskog, K. G. and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a):631–639.

Judi, H. M., Ashaari, N. S., Mohamed, H., and Wook, T. M. T. (2011). Students profile based on attitude towards statistics. *Procedia-Social and Behavioral Sciences*, 18:266–272.

Kamata, A., Bauer, D. J., and Miyazaki, Y. (2008). Multilevel measurement modeling. *Multilevel Modeling of Educational Data*, pages 345–388.

Kamata, A. and Vaughn, B. K. (2011). Multilevel IRT modeling. In Hox, J. and Roberts, J. K., editors, *Handbook of advanced multilevel analysis*, pages 41–57. Routledge, London.

Kang, H. A., Sales, A., and Whittaker, T. A. (2021). Flow with an intelligent tutor: A latent variable modeling approach to tracking flow during artificial tutoring. *PsyArXiv*.

Kankaraš, M., Vermunt, J. K., and Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, 40(2):279–310.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, 44(2):457–467.

Khatun, R. (2019). Rapidly changing globalized economy and its impact on education in the era of globalization. *Research Review International Journal of Multidisciplinary*, 3085:1197–1200.

Kim, E. S., Dedrick, R. F., Cao, C., and Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, 51(6):881–898.

Kim, S. and Lee, W. (2004). *IRT scale linking methods for mixed-format tests*, volume 5. ACT, Incorporated, Iowa City.

Kocdar, S., Karadeniz, A., Bozkurt, A., and Buyuk, K. (2018). Measuring self-regulation in self-paced open and distance learning environments. *International Review of Research in Open and Distributed Learning*, 19(1):25–43.

Koch, T., Holtmann, J., Bohn, J., and Eid, M. (2018). Explaining general and specific factors in longitudinal, multimethod, and bifactor models: Some caveats and recommendations. *Psychological Methods*, 23(3):1–19.

Kohli, N. and Harring, J. R. (2013). Modeling growth in latent variables using a piecewise function. *Multivariate Behavioral Research*, 48(3):370–397.

Kolen, M. J. and Brennan, R. L. (2013). *Test equating: Methods and practices*. Springer Science & Business Media, Berlin.

Krishnasamy, S., Ling, L., and Kim, T. (2020). Improving learning experience of probability and statistics using multimedia system. *International Journal of Emerging Technologies in Learning (iJET)*, 15(1):77–87.

Lambert, S. and Alony, I. (2018). Embedding moocs in academic programmes as a part of curriculum transformation: A pilot case study. In Li, K. C., Yuen, K. S., and Wong, B. T. M., editors, *Innovations in open and flexible education*, pages 73–81. Springer, Singapore.

Lanza, S. T., Tan, X., and Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1):1–26.

Lavidas, K., Barkatsas, T., Manesis, D., and Gialamas, V. (2020). A structural equation model investigating the impact of tertiary students' attitudes toward statistics, perceived competence at mathematics, and engagement on statistics performance. *Statistics Education Research Journal*, 19(2):27–41.

Lawley, D. N. (1940). Vi.—the estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 60(1):64–82.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor analysis as a statistical method*. Butterworths, London.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin, New York.

Legaki, N.-Z., Xi, N., Hamari, J., Karpouzis, K., and Assimakopoulos, V. (2020). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of Human-Computer Studies*, 144:102496.

Lehman, D. R. and Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, 26(6):952.

Lesser, L. and Pearl, D. (2008). Functional fun in statistics teaching: Resources, research and recommendations. *Journal of Statistics Education*, 16(3):1–10.

Lizzio, A. and Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3):263–275.

López Lamezón, S., Rodríguez López, R., Amador Aguilar, L. M., and Azcuy Lorenz, L. M. (2018). Social significance of a virtual environment for the teaching and learning of descriptive Statistics in Medicine degree course. *Humanidades Médicas*, 18(1):50–63.

Lord, F. (1952). *A Theory of Test Scores (Psychometric Monograph No. 7)*. Psychometric Corporation, Richmond.

Lubke, G. and Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41(4):499–532.

Lukočienė, O. and Vermunt, J. K. (2009). Determining the number of components in mixture models for hierarchical data. In Fink, A., Lausen, B., Seidel, W., and Ultsch, A., editors, *Advances in Data Analysis, Data Handling and Business Intelligence*, pages 241–249. Springer, Berlin.

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., and Arendasy, M. (2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *British Journal of Educational Psychology*, 83(4):535–549.

Magidson, J. and Vermunt, J. K. (2004). Latent class models. In Kaplan, D., editor, *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, pages 175–198. SAGE Publications Inc, New York.

Maij-de Meij, A. M., Kelderman, H., and van der Flier, H. (2008). Fitting a mixture item response theory model to personality questionnaire data: Characterizing latent classes and investigating possibilities for improving prediction. *Applied Psychological Measurement*, 32(8):611–631.

Mäkikangas, A., Tolvanen, A., Aunola, K., Feldt, T., Mauno, S., and Kinnunen, U. (2018). Multilevel latent profile analysis with covariates: Identifying job characteristics profiles in hierarchical data as an example. *Organizational Research Methods*, 21(4):931–954.

Marengo, D., Miceli, R., Rosato, R., and Settanni, M. (2018). Placing multiple tests on a common scale using a post-test anchor design: Effects of item position and order on the stability of parameter estimates. *Frontiers in Applied Mathematics and Statistics*, 4:50.

Marsh, H. W. and Grayson, D. (1994). Longitudinal confirmatory factor analysis: common, time-specific, item-specific, and residual-error components of variance. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(2):116–145.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2):149–174.

McCutcheon, A. L. (1987). *Latent class analysis.* Number 64. SAGE Publications Inc, New York.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34(1):100–117.

McMullen, J. and Hickendorff, M. (2018). Latent variable mixture models in research on learning and individual differences. *Learning and Individual Differences*, 66:1–3.

Meredith, W. and Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1):107–122.

Molenaar, I. (2021). Personalisation of learning: Towards hybrid human-AI learning technologies. In Vincent-Lancrin, S., editor, *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, page 57. OECD Publishing.

Molenaar, P. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2):181–202.

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., and Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 29(1):142–163.

Moustaki, I. (2003). A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology*, 56(2):337–357.

Moustaki, I., Jöreskog, K. G., and Mavridis, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of LISREL and IRT approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(4):487–513.

Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65(3):391–411.

Murray, T. and Arroyo, I. (2002). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In Cerri, S. A., Gouardères, G., and Paraguaçu, F., editors, *ITS: International Conference on Intelligent Tutoring Systems*, pages 749–758, Berlin. Springer.

Muslim, A., Chatti, M. A., and Guesmi, M. (2020). Open learning analytics: a systematic literature review and future perspectives. In Muslim, A., Chatti, M. A., and Guesmi, M., editors, *Artificial Intelligence Supported Educational Technologies*, pages 3–29. Springer, Berlin.

Muslim, A., Chatti, M. A., Mahapatra, T., and Schroeder, U. (2016). A rule-based indicator definition tool for personalized learning analytics. In Drachsler, H. and Greller, W., editors, *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 264–273, New York. ACM.

Muthén, B. (1997a). Latent variable growth modeling with multilevel data. In Berkane, M., editor, *Latent variable modeling and applications to causality*, pages 149–161. Springer, New York.

Muthén, B. (1997b). Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology*, 27(1):453–480.

Muthén, B. (2004). Latent variable analysis. In Kaplan, D., editor, *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, pages 345–368. SAGE Publications Inc, New York.

Muthén, B., Brown, C., Khoo, S., Yang, C., and Jo, B. (1998). *General growth mixture modeling of latent trajectory classes: Perspectives and prospects*. Paper presented at the Meeting of the Prevention Science and Methodology Group, Tempe, AZ.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4):557–585.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4):338–354.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29(1):81–117.

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1):1–18.

Nylund-Gibson, K., Grimm, R., Quirk, M., and Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):439–454.

Nylund-Gibson, K., Grimm, R. P., and Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6):967–985.

Nylund-Gibson, K. L., Asparouhov, T., and Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4):535–569.

OECD (2015). *Students, computers and learning: Making the connection.* OECD publishing, Paris.

OECD (2021). *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots.* OECD publishing, Paris.

Ouyang, F. and Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2:100020.

Pacella, D., Fabbricatore, R., Iodice D'Enza, A., Galluccio, C., and Palumbo, F. (2022). Teaching STEM subjects in non-STEM degrees: an Adaptive Learning model for teaching Statistics. In Ouyang, F., Jiao, P., McLaren, B. M., and Alavi, A., editors, *Artificial Intelligence in STEM Education: The Paradigmatic Shifts in Research, Education, and Technology.* Chapman and Hall/CRC, Boca Raton.

Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., et al. (2018). OnTask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics*, 5(3):235–249.

Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement in Education*, 16(3):223–243.

Paviotti, G., Rossi, P. G., and Zarka, D. (2012). *Intelligent Tutoring Systems: An overview.* Pensa Multimedia, Lecce.

Peugh, J. and Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4):616–639.

Pintrich, P. R., Wolters, C. A., and Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In Schraw, G. and Impara, J. C., editors, *Issues in the Measurement of Metacognition.* Buros Institute of Mental Measurements, Lincoln.

Potter, G., Wong, J., Alcaraz, I., Chi, P., et al. (2016). Web application teaching tools for Statistics using R and Shiny. *Technology Innovations in Statistics Education*, 9(1).

Preacher, K. J., Wichman, A. L., MacCallum, R. C., and Briggs, N. E. (2008). *Latent growth curve modeling.* Sage, Thousand Oaks.

Rabe-Hesketh, S. and Skrondal, A. (2008). Classical latent variable models for medical research. *Statistical Methods in Medical Research*, 17(1):5–32.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2):167–190.

Rabin, L., Fink, L., Krishnan, A., Fogel, J., Berman, L., and Bergdoll, R. (2018). A measure of basic math skills for use with undergraduate statistics students: The MACS. *Statistics Education Research Journal*, 17(2):179–195.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Raffalovich, L. E. and Bohrnstedt, G. W. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data. *Sociological Methods & Research*, 15(4):385–405.

Rajiah, K., Coumaravelou, S., and Ying, O. W. (2014). Relationship of test anxiety, psychological distress and academic motivation among first year undergraduate pharmacy students. *International Journal of Applied Psychology*, 4(2):68–72.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika*, 20(2):93–111.

Rasbash, J. and Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19(4):337–350.

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

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage, Thousand Oaks.

Raykov, T. (2007). Longitudinal analysis with regressions among random effects: A latent variable modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(1):146–169.

Raykov, T. and Marcoulides, G. A. (2011). *Introduction to Psychometric Theory.* Routledge, London.

Reckase, M. D. (2009). *Multidimensional Item Response Theory.* Springer-Verlag, Berlin.

Reise, S. P., Moore, T. M., and Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6):544–559.

Reise, S. P., Ventura, J., Nuechterlein, K. H., and Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment*, 84(2):126–136.

Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, 138(2):353.

Richardson, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika*, 1(2):33–49.

Rijmen, F., Tuerlinckx, F., De Boeck, P., and Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8(2):185.

Rosli, M. K. and Maat, S. M. (2017). Attitude towards statistics and performance among post-graduate students. In Puteh, M., Hamid, N. Z. A., and Adenan, N. H., editors, *AIP Conference Proceedings*, volume 1847, page 030004, New York. AIP Publishing LLC.

Ruiz-Mallén, I. and Heras, M. (2020). What sustainability? Higher education institutions' pathways to reach the Agenda 2030 goals. *Sustainability*, 12(4):1290.

Russo, A., Farcomeni, A., Pittau, M. G., and Zelli, R. (2022). Covariate-modulated rectangular latent markov models with an unknown number of regime profiles. *Statistical Modelling*.

Samejima, F. (1969). *Estimation of Ability Using a Response Pattern of Graded Scores*. Educational Testing Service, Princeton.

Samejima, F. (2011). The general graded response model. In Nering, M. and Ostini, R., editors, *Handbook of polytomous item response theory models*, pages 87–118. Routledge, London.

Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics*, 52(2):650–663.

Schmid, J. and Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1):53–61.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3):333–343.

Secore, S. (2017). Social constructivism in online learning: Andragogical influence and the effectual educator. *e-mentor*, 70(3):4–9.

Seery, N., Delahunty, T., Canty, D., and Buckley, J. (2017). Illustrating educational development through ipsative performance in design based education. In de Vries, M., Litowitz, L., and Warner, S., editors, *PATT2017: Technology & Engineering Education – Fostering the Creativity of Youth around the Globe*, pages 1–9. Millersville University of Pennsylvania, Philadelphia.

Sesé, A., Jiménez, R., Montaño, J.-J., and Palmer, A. (2015). Can attitudes towards statistics and statistics anxiety explain students' performance. *Psicodidáct*, 20(2):285–304.

Shin, M., No, U., and Hong, S. (2019). Comparing the robustness of stepwise mixture modeling with continuous nonnormal distal outcomes. *Educational and Psychological Measurement*, 79(6):1156–1183.

Shute, V. J. and Zapata-Rivera, D. (2012). Adaptive educational systems. *Adaptive Technologies for Training and Education*, 7(27):1–35.

Siew, C. S., McCartney, M. J., and Vitevitch, M. S. (2019). Using network science to understand statistics anxiety among college students. *Scholarship of Teaching and Learning in Psychology*, 5(1):75–89.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall/CRC, London.

Skrondal, A. and Rabe-Hesketh, S. (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, 34(4):712–745.

Slavin, R. E. (2019). *Educational psychology: Theory and practice.* Pearson, London.

Smid, S. C., Depaoli, S., and Van De Schoot, R. (2020). Predicting a distal outcome variable from a latent growth model: ML versus Bayesian estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2):169–191.

Song, X., Xia, Y., and Zhu, H. (2017). Hidden Markov latent variable models with multivariate longitudinal data. *Biometrics*, 73(1):313–323.

Sorge, C. (2001). *Impact of engineering students' attitudes on achievement in statistics: A structural model.* PhD thesis, University of New Mexico.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15:201–293.

Spearman, C. (1961). General intelligence, objectively determined and measured. In Jenkins, J. J. and Paterson, D. G., editors, *Studies in individual differences: The search for intelligence*, pages 59–73. Appleton-Century-Crofts, Norwalk.

Stapleton, L. M. (2006). Using multilevel structural equation modeling techniques with complex sample data. In Hancock, G. R. and Mueller, R. O., editors, *Structural Equation Modeling: A Second Course*, pages 345–383. Information Age Publishing, Charlotte.

Stapleton, L. M., McNeish, D. M., and Yang, J. S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, 51(3-4):317–330.

Steel, P., Brothen, T., and Wambach, C. (2001). Procrastination and personality, performance, and mood. *Personality and Individual Differences*, 30(1):95–106.

Stevens, J. P. (2012). *Applied Multivariate Statistics for the Social Sciences.* Routledge, London.

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis? In LeBouton, K. J., editor, *Proceedings of the 31st Annual SAS Users Group International Conference*, pages 200–31. SAS Institute Inc, Cary.

Tay, L., Newman, D. A., and Vermunt, J. K. (2011). Using mixed-measurement item response theory with covariates (MM-IRT-C) to ascertain observed and unobserved measurement equivalence. *Organizational Research Methods*, 14(1):147–176.

te Marvelde, J. M., Glas, C. A., Van Landeghem, G., and Van Damme, J. (2006). Application of multidimensional item response theory models to longitudinal data. *Educational and Psychological Measurement*, 66(1):5–34.

Tempelaar, D. T., Van Der Loeff, S. S., and Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal*, 6(2):78–102.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47(2):175–186.

Thissen, D. and Steinberg, L. (2020). An intellectual history of parametric item response theory models in the twentieth century. *Chinese/English Journal of Educational Measurement and Evaluation*, 1(5):23–39.

Thomas, C. L., Cassady, J. C., and Heller, M. L. (2017). The influence of emotional intelligence, cognitive test anxiety, and coping strategies on undergraduate academic performance. *Learning and Individual Differences*, 55:40–48.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association, Washington.

Thomson, G. H. (1938). *The Factorial Analysis of Human Ability*. University of London Press, London.

Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16(7):433–451.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38(5):406–427.

Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* University of Chicago Press, Chicago.

Toomaneejinda, A. (2017). Zone of proximal development, dynamic assessment and learner empowerment. *LEARN Journal: Language Education and Acquisition Research Network*, 10(1):176–185.

Tremblay, P., Gardner, R., and Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science*, 32(1):40–48.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43(1):39–55.

Tutz, G. (2020). A taxonomy of polytomous item response models. *arXiv preprint*.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3):227–265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221.

Veenman, M. V. (2007). The assessment and instruction of self-regulation in computer-based environments: a discussion. *Metacognition and Learning*, 2(2):177–183.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, 25(3):283–294.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33(1):213–239.

Vermunt, J. K. (2008). Multilevel latent variable modeling: An application in education testing. *Austrian Journal of Statistics*, 37(3&4):285–299.

Vermunt, J. K. (2010a). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4):450–469.

Vermunt, J. K. (2010b). Longitudinal research using mixture models. In van Montfort, K., Oud, J. H. L., and Satorra, A., editors, *Longitudinal Research with Latent Variables*, pages 119–152. Springer, Berlin.

Vermunt, J. K. and Magidson, J. (1922). Structural equation models: Mixture models. *Encyclopedia of Statistics in Behavioral Science*, 1927.

Vermunt, J. K. and Magidson, J. (2016). *Technical guide for Latent GOLD 5.1: Basic, advanced, and syntax.* Statistical Innovations Inc, Belmont.

Vermunt, J. K. and Magidson, J. (2020). *Upgrade manual for Latent GOLD 6.0.* Statistical Innovations Inc., Belmont.

Vermunt, J. K. and Magidson, J. (2021). How to perform three-step latent class analysis in the presence of measurement non-invariance or differential item functioning. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(3):356–364.

Vermunt, J. K., Tran, B., and Magidson, J. (2008). Latent class models in longitudinal research. In Menard, S., editor, *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, pages 373–385. Academic Press, Cambridge.

Von Davier, M., Xu, X., and Carstensen, C. H. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika*, 76(2):318–336.

Vygotskij, L. (1978). Interaction between learning and development. In Gauvain, M. and Cole, M., editors, *Readings on the Development of Children*, pages 29–36. W. H. Freeman and Company, New York.

Walker, E. R. and Brakke, K. E. (2017). Undergraduate psychology students' efficacy and attitudes across introductory and advanced statistics courses. *Scholarship of Teaching and Learning in Psychology*, 3(2):132.

Walkington, C. and Bernacki, M. L. (2019). Personalizing algebra to students' individual interests in an intelligent tutoring system: Moderators of impact. *International Journal of Artificial Intelligence in Education*, 29(1):58–88.

Wang, C., Kohli, N., and Henn, L. (2016). A second-order longitudinal model for binary outcomes: Item response theory versus structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(3):455–465.

Wang, C. and Nydick, S. W. (2020). On longitudinal item response theory models: A didactic. *Journal of Educational and Behavioral Statistics*, 45(3):339–368.

Wang, J. and Wang, X. (2019). *Structural Equation Modeling: Applications Using Mplus*. John Wiley & Sons, Hoboken.

Wang, L., Sy, A., Liu, L., and Piech, C. (2017). Learning to Represent Student Knowledge on Programming Exercises Using Deep Learning. In Hu, X., Barnes, T., Hershkovitz, A., and Paquette, L., editors, *Proceedings of the 10th International Conference on Educational Data Mining*, pages 324–329. ERIC, Washington.

Wang, W., Wilson, M., and Adams, R. J. (1997). Rasch models for multidimensionality between items and within items. In Wilson, M., Englehard, G., and Draney, K., editors, *Objective Measurement: Theory into Practice*, volume 4, pages 139–155. Praeger, Westport.

Wechsler, D. (2003). *Wechsler intelligence scale for children-WISC-IV*. The Psychological Corporation, San Antonio.

Whitney, B. M., Cheng, Y., Brodersen, A. S., and Hong, M. R. (2019). The scale of student engagement in statistics: Development and initial validation. *Journal of Psychoeducational Assessment*, 37(5):553–565.

Woo, S. E., Jebb, A. T., Tay, L., and Parrigon, S. (2018). Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science. *Organizational Research Methods*, 21(4):814–845.

Yang, C. C. (2005). MIMIC latent class analysis model for alcoholic diagnosis. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1):130–147.

Yang, C. C. and Yang, C. C. (2007). Separating latent classes by information criteria. *Journal of Classification*, 24(2):183–203.

Yockey, R. D. (2016). Validation of the short form of the academic procrastination scale. *Psychological Reports*, 118(1):171–179.

Zare, H., Rastegar, A., and Hosseini, S. M. D. (2011). The relation among achievement goals and academic achievement in statistics: the mediating role of statistics anxiety and statistics self-efficacy. *Procedia-Social and Behavioral Sciences*, 30:1166–1172.

Zheng, X. (2017). *Latent Growth Curve Analysis with Item Response Data: Model Specification, Estimation, and Panel Attrition*. PhD thesis, University of Maryland.

Zhu, Y., Steele, F., and Moustaki, I. (2017). A general 3-step maximum likelihood approach to estimate the effects of multiple latent categorical variables on a distal outcome. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5):643–656.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, M., Pintrich, P. R., and Zeidner, M., editors, *Handbook of Self-Regulation*, pages 13–39. Elsevier, Amsterdam.

Zimmerman, B. J. and Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary Educational Psychology*, 39(2):145–155.

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall/CRC, New York.

Zwinderman, A. H. (1991). A generalized rasch model for manifest predictors. *Psychometrika*, 56(4):589–600.

Appendix

A Latent GOLD syntax for the multilevel latent variable model

```
options
maxthreads=8;
   algorithm
      tolerance=1e-08 emtolerance=0,08 emiterations=5000
      nriterations=50;
   startvalues
      seed=7854 sets=16 tolerance=1e-6 iterations=50;
   bayes
      categorical=1 variances=1 latent=1 poisson=1;
   montecarlo
      seed=0 sets=0 replicates=500 tolerance=1e-08;
   quadrature  nodes=10;
   missing  includeall;
   output
      parameters= first  betaopts=wl standarderrors
      profile probmeans=posterior bivariateresiduals
      estimatedvalues=model;
variables
  groupid ID;

  dependent T1_S_ArithmeticMean_A ordinal,
  T1_S_ClassVar_A ordinal, T1_S_GraphForQuant_A ordinal,
  T1_S_Median_A ordinal, T1_S_Percentiles_A ordinal,
  T1_S_StandardDeviation_A ordinal,
  T1_S_SkewnessNKurtosis_A ordinal,
  T1_S_NormalDist_A ordinal,
  T1_S_CumulativeFrequences_A ordinal,
  T1_S_Zscores_A ordinal,
  T1_S_FrequencyTables_J ordinal,
  T1_S_ClassVar_J ordinal,
  T1_S_GraphForQuant_J ordinal, T1_S_Mode_J ordinal,
  T1_S_Percentiles_J ordinal,
```

```
T1_S_StandardDeviation_J ordinal ,
T1_S_SkewnessNKurtosis_J ordinal ,
T1_S_NormalDist_J ordinal , T1_S_Zscores_J ordinal ,
T1_S_FrequencyTables_K ordinal ,
T1_S_ClassVar_K ordinal , T1_S_GraphForCateg_K ordinal ,
T1_S_ArithmeticMean_J ordinal , T1_S_Mode_K ordinal ,
T1_S_ArithmeticMean_K ordinal ,
T1_S_Percentiles_K ordinal ,
T1_S_StandardDeviation_K ordinal ,
T1_S_SkewnessNKurtosis_K ordinal ,
T1_S_NormalDist_K ordinal ,
T1_S_Zscores_K ordinal , T2_S_SamplingSpace_A ordinal ,
T2_S_Probability_A ordinal ,
T2_S_ConditionalProb_A ordinal ,
T2_S_Bernoulli_A ordinal , T2_S_Binomial_A ordinal ,
T2_S_Uniform_A ordinal , T2_S_Gaussian_A ordinal ,
T2_S_SamplingMean_A ordinal ,
T2_S_SamplingVariance_A ordinal ,
T2_S_CentralLimitTheorem_A ordinal ,
T2_S_SamplingSpace_J ordinal ,
T2_S_Probability_J ordinal ,
T2_S_ConditionalProb_J ordinal ,
T2_S_Bernoulli_J ordinal ,
T2_S_Binomial_J ordinal , T2_S_Uniform_J ordinal ,
T2_S_Gaussian_J ordinal , T2_S_SamplingMean_J ordinal ,
T2_S_SamplingVariance_J ordinal ,
T2_S_CentralLimitTheorem_J ordinal ,
T2_S_SampleSpace_K ordinal ,
T2_S_Probability_K ordinal ,
T2_S_ConditionalProb_K ordinal ,
T2_S_Bernoulli_K ordinal , T2_S_Binomial_K ordinal ,
T2_S_UniformBin_K ordinal , T2_S_Gaussian_K ordinal ,
T2_S_SamplingMean_K ordinal ,
T2_S_SamplingVariance_K ordinal ,
T2_S_CentralLimitTheorem_K ordinal ,
T3_S_NullandAlternativeHypo_A ordinal ,
T3_S_TypeITypeIIErrors_A ordinal ,
T3_S_HypoTestPopMeanZTest_A ordinal ,
```

```
T3_S_HypoTestPopMeanTTest_A ordinal,
T3_S_HypoTestDiffPopMeanTTest_A3,
T3_S_HypoTestDiffPopMeanTTest_A5,
T3_S_HypoTestPopProp_A ordinal,
T3_S_HypoTestRatioPopVar_A ordinal,
T3_S_Correlation_A ordinal,
T3_S_EtaSquared_A ordinal,
T3_S_StatSignificance_K ordinal,
T3_S_TypeITypeIIErrors_J ordinal,
T3_S_HypoTestPopMeanZTest_J ordinal,
T3_S_HypoTestPopMeanTTest_J ordinal,
T3_S_HypoTestDiffPopMeanTTest_J3,
T3_S_HypoTestDiffPopMeanTTest_J4,
T3_S_HypoTestPopProp_J ordinal,
T3_S_HypoTestRatioPopVar_J ordinal,
T3_S_Correlation_J ordinal,
T3_S_ChiSquareTest_J ordinal,
T3_S_NullandAlternativeHypo_K ordinal,
T3_S_StatSignificance_J ordinal,
T3_S_HypoTestPopMeanZTest_K2,
T3_S_HypoTestPopMeanZTest_K5,
T3_S_HypoTestPopMeanTTest_K ordinal,
T3_S_HypoTestDiffPopMeanTTest_K ordinal,
T3_S_HypoTestPopProp_K ordinal,
T3_S_HypoTestRatioPopVar_K ordinal,
T3_S_ContingencyTable_K ordinal,
T3_S_EtaSquared_K ordinal;

latent GClass group nominal 3, Application nominal 3,
Judgement nominal 3, Knowledge nominal 3;

independent Time nominal, Sex, PMP_Total, AMS_RAI,
SAS_Examination, SAS_Help, SAS_Interpretation,
SATS_Affect, SATS_Cognitive, SATS_Value,
SATS_Difficulty, MSLQ_SelfEfficacy, MSLQ_TextAnxiety,
MSLQ_CognitiveStrat, MSLQ_SelfRegulation, APS_Total,
ENG_Affective, ENG_Behavioral, ENG_Cognitive;
```

```
equations
  GClass <- 1 + Sex + PMP_Total + AMS_RAI +
  + SAS_Examination + SAS_Help + SAS_Interpretation +
  + SATS_Affect + SATS_Cognitive + SATS_Value +
  + SATS_Difficulty + MSLQ_SelfEfficacy +
  + MSLQ_TextAnxiety + MSLQ_CognitiveStrat +
  + MSLQ_SelfRegulation + APS_Total + ENG_Affective +
  + ENG_Behavioral + ENG_Cognitive;

  Application <- 1|GClass +time;
  Judgement <- 1|GClass +time;
  Knowledge <- 1|GClass+time;

  T3_S_NullandAlternativeHypo_A <- 1 + Application;
  T3_S_TypeITypeIIErrors_A <- 1 + Application;
  T3_S_HypoTestPopMeanZTest_A <- 1 + (1)Application;
  T3_S_HypoTestPopMeanTTest_A <- 1 + Application;
  T3_S_HypoTestDiffPopMeanTTest_A3 <- 1 + Application;
  T3_S_HypoTestDiffPopMeanTTest_A5 <- 1 + Application;
  T3_S_HypoTestPopProp_A <- 1 + Application;
  T3_S_HypoTestRatioPopVar_A <- 1 + Application;
  T3_S_Correlation_A <- 1 + Application;
  T3_S_EtaSquared_A <- 1 + Application;
  T3_S_TypeITypeIIErrors_J <- 1 + Judgement;
  T3_S_HypoTestPopMeanZTest_J <- 1 + (1)Judgement;
  T3_S_HypoTestPopMeanTTest_J <- 1 + Judgement;
  T3_S_HypoTestDiffPopMeanTTest_J3 <- 1 + Judgement;
  T3_S_HypoTestDiffPopMeanTTest_J4 <- 1 + Judgement;
  T3_S_HypoTestPopProp_J <- 1 + Judgement;
  T3_S_HypoTestRatioPopVar_J <- 1 + Judgement;
  T3_S_Correlation_J <- 1 + Judgement;
  T3_S_ChiSquareTest_J <- 1 + Judgement;
  T3_S_NullandAlternativeHypo_K <- 1 + Knowledge;
  T3_S_StatSignificance_J <- 1 + Knowledge;
  T3_S_HypoTestPopMeanZTest_K2 <- 1 + (1)Knowledge;
  T3_S_HypoTestPopMeanZTest_K5 <- 1 + Knowledge;
  T3_S_HypoTestPopMeanTTest_K <- 1 + Knowledge;
  T3_S_HypoTestDiffPopMeanTTest_K <- 1 + Knowledge;
```

```
T3_S_HypoTestPopProp_K <- 1 + Knowledge;
T3_S_HypoTestRatioPopVar_K <- 1 + Knowledge;
T3_S_ContingencyTable_K <- 1 + Knowledge;
T3_S_EtaSquared_K <- 1 + Knowledge;
T3_S_StatSignificance_K <- 1 + Knowledge;
T1_S_ArithmeticMean_A <- 1 + (1) Application;
T1_S_ClassVar_A <- 1 + Application;
T1_S_GraphForQuant_A <- 1 + Application;
T1_S_Median_A <- 1 + Application;
T1_S_Percentiles_A <- 1 + Application;
T1_S_StandardDeviation_A <- 1 + Application;
T1_S_SkewnessNKurtosis_A <- 1 + Application;
T1_S_NormalDist_A <- 1 + Application;
T1_S_CumulativeFrequences_A <- 1 + Application;
T1_S_Zscores_A <- 1 + Application;
T1_S_FrequencyTables_J <- 1 + (1)Judgement;
T1_S_ClassVar_J <- 1 + Judgement;
T1_S_GraphForQuant_J <- 1 + Judgement;
T1_S_Mode_J <- 1 + Judgement;
T1_S_Percentiles_J <- 1 + Judgement;
T1_S_StandardDeviation_J <- 1 + Judgement;
T1_S_SkewnessNKurtosis_J <- 1 + Judgement;
T1_S_NormalDist_J <- 1 + Judgement;
T1_S_Zscores_J <- 1 + Judgement;
T1_S_ArithmeticMean_J <- 1 + Judgement;
T1_S_FrequencyTables_K <- 1 + (1)Knowledge;
T1_S_ClassVar_K <- 1 + Knowledge;
T1_S_GraphForCateg_K <- 1 + Knowledge;
T1_S_Mode_K <- 1 + Knowledge;
T1_S_ArithmeticMean_K <- 1 + Knowledge;
T1_S_Percentiles_K <- 1 + Knowledge;
T1_S_StandardDeviation_K <- 1 + Knowledge;
T1_S_SkewnessNKurtosis_K <- 1 + Knowledge;
T1_S_NormalDist_K <- 1 + Knowledge;
T1_S_Zscores_K <- 1 + Knowledge;
T2_S_SamplingSpace_A <- 1 + Application;
T2_S_Probability_A <- 1 + Application;
T2_S_ConditionalProb_A <- 1 + Application;
```

```
T2_S_Bernoulli_A <- 1 + Application;
T2_S_Binomial_A <- 1 + Application;
T2_S_Uniform_A <- 1 + Application;
T2_S_Gaussian_A <- 1 + Application;
T2_S_SamplingMean_A <- 1 + (1) Application;
T2_S_SamplingVariance_A <- 1 + Application;
T2_S_CentralLimitTheorem_A <- 1 + Application;
T2_S_SamplingSpace_J <- 1 + Judgement;
T2_S_Probability_J <- 1 + Judgement;
T2_S_ConditionalProb_J <- 1 + Judgement;
T2_S_Bernoulli_J <- 1 + Judgement;
T2_S_Binomial_J <- 1 + Judgement;
T2_S_Uniform_J <- 1 + Judgement;
T2_S_Gaussian_J <- 1 + (1)Judgement;
T2_S_SamplingMean_J <- 1 + Judgement;
T2_S_SamplingVariance_J <- 1 + Judgement;
T2_S_CentralLimitTheorem_J <- 1 + Judgement;
T2_S_SampleSpace_K <- 1 + Knowledge;
T2_S_Probability_K <- 1 + Knowledge;
T2_S_ConditionalProb_K <- 1 + (1)Knowledge;
T2_S_Bernoulli_K <- 1 + Knowledge;
T2_S_Binomial_K <- 1 + Knowledge;
T2_S_UniformBin_K <- 1 + Knowledge;
T2_S_Gaussian_K <- 1 + Knowledge;
T2_S_SamplingMean_K <- 1 + Knowledge;
T2_S_SamplingVariance_K <- 1 + Knowledge;
T2_S_CentralLimitTheorem_K <- 1 + Knowledge;
```

B R Code for the three-step rectangular latent Markov model

Step 1: Multidimensional latent class IRT model

```
library(readxl)
library(MultiLCIRT)
library(dplyr)
library(mclust)
library(mclogit)

# Read the data file
Data = read_excel("data_matrix.xlsx", na = "999")

# Select items of Time 1 and order them according to the
    considered dimensions
Data_lc_r = dplyr::select(Data, starts_with('T1_S_'))
K = dplyr::select(Data_lc_r, ends_with('_K'))
A = dplyr::select(Data_lc_r, ends_with('_A'))
J = dplyr::select(Data_lc_r, ends_with('_J'))
Data_lc_r = cbind(K, A, J)

# Define the matrix with item indices according to the
    measured dimensions (for each dimension, the first item is
    the reference for model identifiability)
multi = rbind(c(3, 1, 2, rep(4:10)), c(14, rep(11:13),
        rep(15:20)), c(23, 21, 22, rep(24:30)))

# Model selection (compare models with different number of
    classes) following the GPCM
GPCM = list()
for (i in 1:5) {
  GPCM[[i]] <- est_multi_poly(Data_lc_r, k = i, link = 2 ,
             disc = 1, difl = 0, output = T, multi = multi)
}

BIC_value = c(GPCM[[1]]$bic, GPCM[[2]]$bic, GPCM[[3]]$bic,
             GPCM[[4]]$bic, GPCM[[5]]$bic)

# Best model according to the BIC
GPCM3 = GPCM[[3]]
```

```
33 # Model parameters
34 GPCM3$piv # Class weights
35 GPCM3$Th # Matrix of support points
36 GPCM3$Bec # Item difficulty parameters
37 GPCM3$gac # Item discriminating parameters
38 GPCM3$Pp # Matrix of posterior probabilities
39
40 # Repeat lines 10-38 for Time 2 and Time 3 and obtain: "Data_lc_r2" and "Data_lc_r3" (datasets); "GPCM3_t2" and "GPCM2_t3" (MultiLCIRT model output)
```

Step 2: Modal class assignment and classification error

```
41 # Classify the observations according to the posterior class probabilities
42 Data_lc_r$Id = rep(1:nrow(Data_lc_r))
43 Data_lc_r$Clus1 = data.frame(rep(0, nrow(Data_lc_r)))
44 for (j in 1:nrow(Data_lc_r)) {
45   Data_lc_r$Clus1[j,] = which.max(GPCM3$Pp[j,])
46 }
47
48 # Repeat lines 41-46 for Time 2 and Time 3 and obtain: "Data_lc_r2$Clus2" and "Data_lc_r3$Clus3"
49
50 # Create a matrix n x T with class assignments
51 total_class = as.data.frame(left_join(Data_lc_r[, c("Id",
52     "Clus1")], Data_lc_r2[, c("Id", "Clus2")], by = c("Id")))
53 total_class = as.matrix(left_join(total_class[, c("Id", "Clus1",
54     "Clus2")], Data_lc_r3[, c("Id", "Clus3")], by = c("Id")))
55
56 # Create a function for the modal D matrix computation
57 Dmatrix = function(outmodel){
58
59 classweight = outmodel$piv
60 numobservation = nrow(outmodel$Pp)
61 numofclasses = length(outmodel$piv)
62 posteriorprob = as.matrix(outmodel$Pp)
63
64 W = posteriorprob == outer(apply(posteriorprob,1, max),
65     rep(1,numofclasses))
66 num = (t(posteriorprob) %*% W)/numobservation
```

```
Dmatrix = num/classweight

return(Dmatrix)
}

# Calculate the D matrix for each Time and store the results in the cD array
cD = array(NA, c(3,3,3))
    cD[,,1] = Dmatrix(GPCM3)
    cD[,,2] = Dmatrix(GPCM3_t2)
    cD[1:2,1:2,3] = Dmatrix(GPCM2_t3)
```

Step 3: BCH correction to account for covariate effect

```
# Cross-tabulate class assignments from previous steps to obtain initial and transitions as if the assignments were realizations of an observed Markov chain
inistart = table(total_class[,1+1], useNA = "always")/sum(table(total_class[,1+1], useNA = "always"))
PI2 = table(total_class[,1+1],total_class[,1+2], useNA = "always")/rowSums(table(total_class[,1+1],total_class[,1+2], useNA = "always"))
PI3 = table(total_class[,1+2],total_class[,1+3], useNA = "always")/rowSums(table(total_class[,1+2],total_class[,1+3], useNA = "always")) # Dropout class for NAs at Time 2 and Time 3

total_class_recod = total_class[,-1]
N = dim(total_class_recod)[1]
for(t in 1:3){
  total_class_recod[is.na(total_class_recod[,t]),t] = max(total_class_recod[,t],na.rm=T)+1
}

modal_class1 = mclust::unmap(total_class_recod[,1])
modal_class2 = mclust::unmap(total_class_recod[,2])
modal_class3 = mclust::unmap(total_class_recod[,3])

# Create "individual" transitions
iK = c(3,3,2) # Number of latent classes for each time point
PI2_dep = array(0,c(N,(iK[2]+1),iK[1]))
for(n in 1:N){
```

```
96   PI2_dep[n,,] = t((modal_class1[n,])%*%t(modal_class2[n,]))
97 }
98
99 PI3_dep = array(0,c(N,(iK[3]+1),iK[2]+1))
100 for(n in 1:N){
101   PI3_dep[n,,] = t((modal_class2[n,])%*%t(modal_class3[n,]))
102 }
103
104 # Create BCH weights from classification error probabilities
105 cDexp2 = diag(4)
106 cDexp2[1:3,1:3] = cD[,,2]
107 cDexp3 = diag(3)
108 cDexp3[1:2,1:2] = cD[1:2,1:2,3]
109
110 wei1 = diag(solve(cD[,,1]))[total_class_recod[,1]]
111 wei2 = diag(solve(cDexp2))[total_class_recod[,2]]
112 wei3 = diag(solve(cDexp3))[total_class_recod[,3]]
113
114 # Select the covariates from the dataset
115 covar = dplyr::select(Data, c("Sex", "PMP_Total",
116     "SAS_Examination", "SAS_Interpretation", "SATS_Affect",
117     "SATS_Difficulty", "MSLQ_SelfEfficacy", "ENG"))
118
119 # Estimate covariate effect at Time 1
120 df_t1 = data.frame(y = factor(mclust::map(modal_class1)),
    covar)
121 out_t1 = mblogit(y ~ ., data = df_t1, weights = wei1)
122
123 # Estimate covariate effect on transitions at Time 2
124 # Starting in state 1 (first row of transition matrix), first
    (arrival) state as reference
125 df_t2_s1 = data.frame(y = factor(mclust::map(PI2_dep[,,1])),
    covar)
126 out_t2_s1 = mblogit(y ~ ., data = df_t2_s1, weights = wei2)
127
128 # Repeat lines 123-126 for each sub-sample of individuals
    defined by the latent classes emerged at Time t considering
     the corresponding classification at Time t+1 as dependent
    variables to estimate the covariate effect on all the other
     transitions at Time 2 and Time 3
```

C Pseudocode of the ML correction in the three-step rectangular latent Markov model

Algorithm 1 Scaling the likelihood computation

function MixtDensityScale($\boldsymbol{\pi}^{(t)}, \log(\boldsymbol{r}_s^{(t)}), I_t$)
 $\boldsymbol{\nu}^{(t)} = \log(\boldsymbol{\pi}^{(t)}) + \log(\boldsymbol{r}_s^{(t)}) = \log(\boldsymbol{\alpha}^{(t)})$
 $scale = \max(\boldsymbol{\nu}^{(t)})$ ▷ Scale factor
 $\boldsymbol{\nu}_{scaled}^{(t)} = \boldsymbol{\nu}^{(t)} - scale$
 $L = \mathbf{1}'[\exp(\boldsymbol{\nu}_{scaled}^{(t)})]$ ▷ Scaled likelihood
 $l = \log(L) + scale$ ▷ Log-likelihood
 return l
end function

Algorithm 2 Forward Filtering, Backward smoothing

function FFBS_Rectangular($\log(\boldsymbol{R}_s), \boldsymbol{h}'_s, \boldsymbol{\pi}^{(1)}, \boldsymbol{\Gamma}, \boldsymbol{I}$)
 ▷ Forward probabilities computation
 $\boldsymbol{\nu}^{(1)} = \log(\boldsymbol{\pi}^{(1)}) + \log(\boldsymbol{r}_s^{(1)}) = \log(\boldsymbol{\alpha}^{(1)})$
 $lscale$ = `MixtDensityScale`($\boldsymbol{\pi}^{(1)}, \log(\boldsymbol{r}_s^{(t)}), I_1$) ▷ Log of the sum of $\boldsymbol{\alpha}^{(1)}$'s elements
 $\boldsymbol{\nu}_{scaled}^{(1)} = \boldsymbol{\nu}^{(1)} - lscale$ ▷ Working parameter
 $\boldsymbol{L}_{\boldsymbol{\alpha}} \leftarrow$ matrix ($I_{max} \times T$) containing the vectors $\log(\boldsymbol{\alpha}^{(t)})$
 $\boldsymbol{L}_{\boldsymbol{\alpha}}[1 : I_1, 1] \leftarrow \log(\boldsymbol{\alpha}^{(1)})$
 for t in $2 : T$ **do**
  if $\boldsymbol{h}'_s[t] == 1$ **then**
   $\boldsymbol{z} = \mathbf{0}_{(I_t,1)}$
   for j in $1 : I_t$ **do**
    $z[j]$ = `MixtDensityScale`($\boldsymbol{\Gamma}[1 : I_{t-1}, j, t-1], \boldsymbol{\nu}_{scaled}^{(t-1)}, I_{t-1}$)
   end for
   $\boldsymbol{\nu}^{(t)} = \boldsymbol{z} + \log(\boldsymbol{r}_s^{(t)})$ ▷ Natural parameter
   $lscale = lscale$ + `MixtDensityScale`($\exp(\boldsymbol{z}), \log(\boldsymbol{r}_s^{(t)}), I_t$) ▷ Update $lscale$
   $\boldsymbol{\nu}_{scaled}^{(t)} = \boldsymbol{\nu}^{(t)} - lscale$ ▷ Update $\boldsymbol{\nu}_{scaled}^{(t)}$
   $\log(\boldsymbol{\alpha}^{(t)}) = \boldsymbol{\nu}^{(t)}$

$\boldsymbol{L_\alpha}[1:I_t,t] \leftarrow \log(\boldsymbol{\alpha}^{(t)})$

end if

end for

$l^{(T)} = lscale^{(T)} = \log(\boldsymbol{\alpha}'^{(T)}\mathbf{1})$ ▷ Log-likelihood of the model

▷ Backward probabilities computation

$\boldsymbol{L_\beta} \leftarrow$ matrix $(I_{max} \times T)$ containing the vectors $\log(\boldsymbol{\beta}^{(t)})$

$\boldsymbol{L_\beta}[1:I_T,T] \leftarrow \log(\boldsymbol{\beta})^{(T)} = \mathbf{0}_{(I_T,1)}$

$\boldsymbol{\nu}^{(T)} \leftarrow$ vector of length T with elements $\log(1/I_T)$

$lscale = \log(I_T)$

for t in $(T-1):1$ **do**

$\boldsymbol{u} = \mathbf{1}_{(I_t,1)}$

if $\boldsymbol{h}_s'[t] == 1$ **then**

$\boldsymbol{z} = \log(\boldsymbol{r}_s^{(t+1)}) + \boldsymbol{\nu}^{(t+1)}$ ▷ Working parameter

$\boldsymbol{\nu}^{(t)} \leftarrow \mathbf{0}_{(I_t,1)}$

for j in $1:I_t$ **do**

$\boldsymbol{\nu}^{(t)}[j] =$ `MixtDensityScale`$(\boldsymbol{\Gamma}[j,1:I_{t+1},t]',\boldsymbol{z},I_{t+1})$

end for

$\log(\boldsymbol{\beta}^{(t)}) = \boldsymbol{\nu}^{(t)} + lscale$ ▷ Natural parameter

$\boldsymbol{L_\beta}[1:I_t,t] \leftarrow \log(\boldsymbol{\beta}^{(t)})$

$\boldsymbol{\nu}^{(t)} = \boldsymbol{\nu}^{(t)} -$ `MixtDensityScale`$(\boldsymbol{u},\boldsymbol{\nu}^{(t)},I_t)$ ▷ Update $\boldsymbol{\nu}^{(t)}$

$lscale = lscale +$ `MixtDensityScale`$(\boldsymbol{u},\boldsymbol{\nu}^{(t)},I_t)$ ▷ Update $lscale$

end if

end for

return $l^{(T)}$, $\boldsymbol{L_\alpha}$, $\boldsymbol{L_\beta}$

end function

Algorithm 3 Baum-Welch algorithm

function RLM_FIXED$(d\mathbf{W},\mathbf{W},\boldsymbol{H},\boldsymbol{I},c\boldsymbol{D},\text{tol}=1e^{-8},\text{maxit}=1000)$

$N \leftarrow$ number of rows in $\mathbf{W}$

$T \leftarrow$ number of columns in $\mathbf{W}$

$I_{max} \leftarrow$ maximum of the vector $\boldsymbol{I}$

$\boldsymbol{\pi}^{(1)} \leftarrow$ matrix $I_{max} \times I_{max}$

$\boldsymbol{\Gamma} \leftarrow$ array $I_{max} \times I_{max} \times (T-1)$

$\boldsymbol{\pi}^{(1)} \leftarrow$ mean for columns of $d\mathbf{W}[[1]]$ ▷ Initialize $\boldsymbol{\pi}^{(1)}$

for t in $2:T$ **do**

$\boldsymbol{\Gamma}[1:I_{t-1},1:I_t,t-1] \leftarrow$ contingency table between $\mathbf{W}[,t-1]$ and $\mathbf{W}[,t]$ divided by row sum ▷ Initialize $\boldsymbol{\Gamma}$

end for

▷ E-Step
$\boldsymbol{Q}_\alpha$, $\boldsymbol{Q}_\beta \leftarrow$ array $N \times T \times I_{max}$
$\boldsymbol{R} \leftarrow$ array $N \times I_{max} \times T$
liks $\leftarrow$ vector $N \times 1$
for t in $1 : T$ **do**
  for s in $1 : I_t$ **do**
    $N^{(t)} \leftarrow$ number of observation at time t
    $\boldsymbol{R}[\boldsymbol{H}[, t] == 1, s, t] \leftarrow \log\{\text{Multinomial}(x = d\mathbf{W}, size = 1, prob = c\boldsymbol{D}[s, , t])\}$ ▷ Evaluating the density of available (non-missing) units
  end for
end for
for i in $1 : N$ **do**
  out = FFBS_Rectangular$(\boldsymbol{R}[i, ,]', \boldsymbol{H}[i,], \boldsymbol{\pi}^{(1)}, \boldsymbol{\Gamma}, \boldsymbol{I})$
  $\boldsymbol{Q}_\alpha[i, ,] \leftarrow \boldsymbol{L}_\alpha$ from out (transposed)
  $\boldsymbol{Q}_\beta[i, ,] \leftarrow \boldsymbol{L}_\beta$ from out (transposed)
  liks$[i] \leftarrow l^{(T)}$ from out
end for
lik $\leftarrow$ sum of the element in liks
$\boldsymbol{V} = \boldsymbol{Q}_\alpha + \boldsymbol{Q}_\beta$
for j in $1 : I_{max}$ **do**
  $\boldsymbol{V}[, , j] \leftarrow \boldsymbol{V}[, , j]$ scaled dividing by the log-likelihood ▷ Compute $\widehat{u}_{si}^{(1)}$
end for
$\boldsymbol{V} = \exp(\boldsymbol{V})$
for t in $1 : T$ **do**
  $\boldsymbol{V}[, t,] \leftarrow \boldsymbol{V}[, t,]$ scaled dividing by row sum
end for
$\boldsymbol{Z} \leftarrow$ array $N \times I_{max} \times I_{max} \times (T - 1)$
for t in $2 : T$ **do**
  liks_temp $\leftarrow$ logarithm of the sum of $\exp\{\boldsymbol{Q}_\alpha + \boldsymbol{Q}_\beta\}$ avoiding underflow
  for c in $1 : I_{t-1}$ **do**
    for d in $1 : I_t$ **do**
      $\boldsymbol{Z}[\boldsymbol{H}[, t] == 1, c, d, t - 1] = \exp\{\log(\boldsymbol{\Gamma}[c, d, t - 1]) + \boldsymbol{Q}_\alpha[\boldsymbol{H}[, t] == 1, t - 1, c] + \boldsymbol{Q}_\beta[\boldsymbol{H}[, t] == 1, t, d] + \boldsymbol{R}[\boldsymbol{H}[, t] == 1, d, t] - \text{liks\_temp}\}$ ▷ Compute $\widehat{v}_{sil}^{(t)}$
    end for
  end for
  $\boldsymbol{Z}[, 1 : I_{t-1}, 1 : I_t, t - 1] \leftarrow \boldsymbol{Z}[, 1 : I_{t-1}, 1 : I_t, t - 1]$ scaled dividing by row sum
end for
lkold $\leftarrow$ lik$*2$
iters = 1

```
48:     while lik-lkold>tol and iters<maxit do
49:         ▷ M-Step
50:         if I_1 > 1 then
51:             π^(1) ← sum for columns of elements in V[, 1, 1 : I_1]
52:             π^(1) ← π^(1) divided by the sum of its elements        ▷ ML estimate of π^(1)
53:         end if
54:         for t in 1 : (T − 1) do
55:             for c in 1 : I_t do
56:                 for d in 1 : I_{t+1} do
57:                     Γ[c, d, t] ← sum of elements in Z[, c, d, t]
58:                     if Γ[c, d, t] < 1e^{−5} then
59:                         Γ[c, d, t] = 1e^{−5}
60:                     end if
61:                 end for
62:                 Γ[c, , t] ← Γ[c, , t] divided by the sum of elements   ▷ ML estimate of Γ
63:             end for
64:         end for
65:         llkseries[iters]=lik
66:         lkold=lik
67:         ▷ Repeat lines 21-45 to perform the E-Step
68:         iters=iters+1
69:     end while
70:     np ← 1′[2 ∗ I] + I_1 − 1
71:     for j in 1 : I_max do
72:         for h in 1 : I_max do
73:             trans = I[−T] == j and I[−1] == h
74:             if any trans then
75:                 np = np + j ∗ (h − 1)                            ▷ Number of parameters
76:             end if
77:         end for
78:     end for
79:     AIC = −2∗lik+2 ∗ np
80:     BIC = −2∗lik+log(N) ∗ np
81:     llkseries = llkseries[1 : (iters − 1)]
82:     return π^(1), Γ, llkseries, lik, AIC, BIC
83: end function
```