Università degli Studi di Napoli Federico II
Scuola delle Scienze Umane e Sociali

# Stat.Edu'21
# New Perspectives in Statistics Education

Proceedings of the International Conference Stat.Edu'21
March 25-26, 2021
Naples, Italy

**Editors**
Cristina Davino, Rosa Fabbricatore, Anna Parola

Federico II University Press

fedOA Press

Università degli Studi di Napoli Federico II

Scuola delle Scienze Umane e Sociali

# Stat.Edu'21

# New Perspectives in Statistics Education

Proceedings of the International Conference Stat.Edu'21
March 25-26, 2021
Naples, Italy

**Editors**

Cristina Davino, Rosa Fabbricatore, Anna Parola

Federico II University Press

fedOA Press

Stat.Edu’21 : New Perspectives in Statistics Education : Proceedings of the International Conference Stat.Edu’21 : March 25-26, 2021 Naples, Italy / editors Cristina Davino, Rosa Fabbricatore, Anna Parola. – Napoli : FedOAPress, 2022. – 84 p. : ill. ; 24 cm.

Accesso alla versione elettronica:
http://www.fedoabooks.unina.it

ISBN: 978-88-6887-124-6
DOI: 10.6093/978-88-6887-124-6

© 2022 FedOAPress - Federico II University Press

Università degli Studi di Napoli Federico II
Centro di Ateneo per le Biblioteche “Roberto Pettorino”
Piazza Bellini 59-60
80138 Napoli, Italy
http://www.fedoapress.unina.it/

Published in Italy

Gli E-Book di FedOAPress sono pubblicati con licenza
Creative Commons Attribution 4.0 International

# Contents

# Preface

The volume collects a selection of papers presented during the Conference "Stat.Edu'21 - New Perspectives in Statistics Education". The conference was held at the Department of Political Sciences of the University of Naples Federico II (25 - 26 March 2021).

The conference was the final event of the "ALEAS - Adaptive LEArning in Statistics", an ERASMUS+ project (`https://aleas-project.eu`) developed in the period $2018 - 2021$. The goal of the project was to design and implement an adaptive learning system able to offer personalised learning paths to students, with the purpose to provide them remedial advice to deal with the "Statistics anxiety". The project was coordinated by Francesco Palumbo from the University of Naples Federico II (Italy) and involved the following partners: Jacobs University Bremen, Germany (Adalbert Wilhelm); Democritus University of Thrace, Greece (Angelos Markos); Polytechnic University of Valencia, Spain (Rosa María Baños); Smarted, Italy (Raffaele Di Fuccio).

Stat.Edu'21 aimed to stimulate discussions, solicitations and contributions around the central theme of ALEAS, the development of adaptive learning systems in the field of Higher Education as complementary tools for traditional courses and promoting a community of practice in this field. The thematic sessions addressed the following themes: Adaptive learning systems, Learning analytics, Self-learning and innovative technologies, Statistical methods and models for assessing learners' ability, Statistical methods for adaptive systems, Students' attitudes towards Statistics assessment, Technology-based platforms in higher education.

The conference provided four sessions: Statistical methods and models for assessing learners' ability; New perspectives in statistical anxiety; Non-conventional learning environments; E-learning in higher education, for a total of 16 presentations. The conference featured five distinguished Keynote Speakers: Katharina Schüller (CEO of Stat-up Statistical Consulting & Data Science; Germany), Davide Taibi (ITD-CNR; Italy), Vincenzo Esposito Vinzi (ESSEC Business School; France), Giuseppe Riva (Catholic University of the Sacred Heart; Italy) and Ruth Kerr (Federica Web Learning; Italy). During the two days PhD, students, researchers, professors from several organisations and relevant stakeholders participated as the audience.

This volume collects twelve papers, a selection of the papers presented to the conference. We are indebted to all the Conference Organizers and Program Committee members for their soliciting presentation of novel and highly valued contributions. Two anonymous referees reviewed each paper to ensure the highest quality of the contributions in the present volume. Therefore, our warmest and sincere thanks for carefully reading the papers go to all the anonymous referees.

The papers report reflections and quantitative studies covering mainly three topics: the assessment of the effects of anxiety or, more generally, of a different attitude in the study of Statistics, tools and methods for assessing training paths and technology-based learning experiences.

*Section 1. Assessment of the effects of a different attitude in the study of Statistics*

The first paper titled "Anxiety towards statistics: Wisdom, Interest and Cognitive Competence as predictors" by María Folgado-Alufre, Marta Miragall, Marian Serrano-Mendizábal, Lucía García-Ubiedod, and Rosa M. Baños proposes a study aiming at analysing the predictor role of the psychological or character strengths with respect to statistical anxiety. The study was conducted on a sample of social and health science Spanish students, and a serial multiple mediator analysis was performed to analyse the data collected through the well-known SAS - Statistical Anxiety Scale.

The second paper, entitled "Statistical Anxiety in undergraduate students:

PLS-PM approach for exploring the role of situational and dispositional factors" by Rosa Fabbricatore, Anna Parola, Giuliana Pepicelli and Francesco Palumbo, focuses on the role of situational and dispositional antecedents in predicting statistical anxiety. The study was conducted on a sample of Italian psychology students and a partial least squared path modelling (PLS-PM) was performed. Also in this study, for data collection the SAS - Statistical Anxiety Scale was used.

The third paper is titled "Understanding Students' Attitude towards Statistics" by Irasianty Frost, and presents the mechanisms of how attitudes toward Statistics change. The study was conducted on a sample of undergraduate students in Economics, and Business Psychology in Berlin. Bayesian Networks were used to analyse the data collected through the SATS-36 - Survey of Attitudes Toward Statistics.

*Section 2. Tools and methods for the assessment of training paths*

The first paper, titled "A general nonparametric classification approach for cognitive diagnostic assessment in introductory statistics" by Evripidis Themelis and Angelos Markos, proposes a non-parametric cognitive diagnostical algorithm that improves upon the General Non-Parametric Classification (GNPC) method. This study showed that a modified version of the GNPC method allows estimating which skills each student has mastered or failed to master in a Statistics course.

The second paper is titled "Challenging in teaching biostatistics in an e-learning environment. The experience of a postgraduate course" by Claudia Franceschini, Danila Azzolina, Ileana Baldi, Giulia Lorenzoni, Corrado Lanera, Paola Berchialla, and Dario Gregori describes the experience of distance learning of two postgraduate teaching courses in Statistics delivered by the University of Padua. The study focuses on the students' satisfaction, which is useful as a starting point for redefining the educational setting for future courses.

The third paper, titled "Content validity assessment of ALEAS statistical reasoning items via think-aloud interviews" by Aikaterini Tsaousi and Angelos Markos, investigates the content validity of a set of ALEAS items for making judgments using think-aloud interviews. Fifteen draft questions were administered to Greek undergraduate students of primary education who had recently completed an introductory statistics course.

The fourth paper is titled “Statistics in Medicine in the degree courses in Nursing in DAD: first results on the experience in Piedmont” by Ilaria Stura, Alessandra Alemanni, and Giuseppe Migliaretti. It focuses on the effect of distance learning on the performance of students of a Medical Statistics course of the Nursing degree in three campus of the University of Turin using logistic regression models.

The fifth paper is titled “Differential performance on OECD PISA tests: Evaluating Item Functioning methods using Mathematics Assessments” by Clelia Cascella and Rosa Fabbricatore. It proposes a study aiming to analyse the hypothesis of unidimensionality of the OECD-PISA mathematics achievement test administered in 2018 through the analysis of both manifest and latent Differential Item Functioning.

*Section 3. Technology-based learning experiences*

The first paper titled “Investigating the impact of Covid-19 related stress and anxiety on distance learning perception” by Maria Iannario, Alfonso Iodice D’Enza, and Rosaria Romano investigates how stress and career-related anxiety impact the perception of distance learning. Data are collected from a survey on Italian university students through the Future career anxiety scale and analysed using exploratory factor analysis and partial least squares-based structural equation modelling.

The second paper titled “Blended & Data-Driven Learning: a teaching innovation project” by Emilia López-Iñesta, Maria T. Sanz, Daniel Garcia-Costa, and Francisco Grimaldo proposes a teaching project framed in the Learning Analytics area. The paper describes the teaching experience adapted to the COVID-19 scenario with pre-service primary teachers of Valencia and the results of Moodle data analysis about the different tools that were used to monitor and evaluate the students.

The third paper is titled "STop Obesity Platform: a gamified learning system to build a new healthier life-style" by Monica Zampella, Eliana Brunetti, Marianna Mugione, Raffaele Di Fuccio, Michela Ponticorvo and Fabrizio Ferrara. It presents the project STOP (STop Obesity Platform, Horizon 2020 Research and Innovation program). The authors discuss the STOP platform as a learning system aimed at facing the challenge of obesity through a digital methodology.

The fourth paper, titled "ALEAS: An interactive application for ubiquitous learning in higher education in statistics" by Raffaele Di Fuccio, Fabrizio Ferrara, Giovanni Siano, and Andrea Di Ferdinando, shows the technical infrastructure of the ALEAS app. ALEAS app aimed to support students' learning by providing several resources, as tests, videos, cartoons, texts, usable through mobile devices.

Naples, November 2021

*Cristina Davino*
*Rosa Fabbricatore*
*Anna Parola*
(University of Naples Federico II, Italy)

# Section 1
# Assessment of the effects of a different attitude in the study of Statistics

# Anxiety towards statistics: Wisdom, Interest and Cognitive Competence as predictors

María Folgado-Alufre*, Marta Miragall**, Marian Serrano-Mendizábal***
Lucía García-Ubiedod****, Rosa M. Baños*****

*Abstract:* Statistical Anxiety (SA) can be *conceptualised* as a negative state of emotional arousal and concern experienced by individuals when dealing with statistics. The aim of this study was to determine the predictors of SA. Specifically, the effect of wisdom -a character strength that helps in acquiring and using knowledge- on SA, and the mediator role of the individual's interest and cognitive competence in statistics, were tested. The sample was composed of 197 Health and Social Science undergraduate Spanish students (56.7% females; age; *M* = 20.87, *SD* = 2.71). They answered several self-reported online questionnaires. A serial multiple mediation showed a significant indirect effect of wisdom on SA through interest and cognitive competence in statistics. Findings highlight the need of increasing interest and competence in students to reduce SA.

*Keywords:* Statistical anxiety, Interest, Cognitive competence.

## *1. 1. Introduction*

The importance of statistics-related knowledge and skills within Social and Health Sciences relies not only on achieving desirable academic outcomes, but also on being able to display proficiency in a professionally useful skillset. However, anxiety toward statistics is one of the main barriers to developing

* Instituto Polibienestar, University of Valencia, maria.folgado@uv.es

**Department of Personality, Evaluation, and Psychological Treatments, University of Valencia; CIBER Fisiopatología Obesidad y Nutrición (CIBEROBN), Instituto Carlos III, Madrid, marta.miragall@uv.es

***Department of Developmental and Educational Psychology, University of Valencia; ERI-Lectura, University of Valencia, m.angeles.serrano@uv.es

****Department of Developmental and Educational Psychology, University of Valencia, lugaru@alumni.uv.es

*****Instituto Polibienestar, University of Valencia; Department of Personality, Evaluation, and Psychological Treatments, University of Valencia; CIBER Fisiopatología Obesidad y Nutrición (CIBEROBN), Instituto Carlos III, Madrid, rosa.banos@uv.es

those skills. Statistical Anxiety (SA) can be conceptualised as a negative state of emotional arousal and concern experienced by individuals when dealing with statistics (Zeidner, 1991). SA is a pervasive problem in the context of university studies, especially in Health and Social Science degrees, such as Psychology, Education, or Sociology (Ruggeri *et al.*, 2008).

Previous research has identified situational, dispositional, and cognitive antecedents of SA. Regarding the dispositional factors, a recent systematic review shows the importance of the type of motivation and attitudes, as well as behaviours and individual characteristics in relation to SA (Cui *et al.*, 2019). In this regard, positive attitudes toward statistics is one of the main predictors of student performance (Hood *et al.*, 2012) that alleviates the adverse effects of SA (Najmi *et al.*, 2018). The most prominent attitude towards statistics is the interest, which provides an incentive for positive learning-directed behaviours. Accordingly, students interested in statistics tend to experience less SA (Macher *et al.*, 2013). Slootmaeckers *et al.* (2014) also point out that interest in statistics enhances students' acquisition of quantitative skills as well as retention of statistics knowledge. Moreover, the knowledge and intellectual skills, together with the perception of cognitive competence, seem to play a key role in alleviating SA (González *et al.*, 2016; Macher *et al.*, 2013).

Related to the dispositional variables, several psychological or character strengths (CS) may act as protective factors of experiencing SA. CS are associated with coping with adversity and difficulties in many different contexts, such as education (Duckworth *et al.*, 2009). Linley (2008) defines CS as "pre-existing capacities for a particular way of acting, thinking or feeling that are authentic and energizing for the individual, and enable optimal functioning, development and performance" (p. 9). From this perspective, CSs constitute the processes or mechanisms that define the virtues. Specifically, the virtue of "wisdom" refers to five CS that entails the acquisition and use of knowledge, such as creativity, curiosity, open-mindedness, love of learning, or taking different perspectives (Peterson & Seligman, 2004).

Despite the key role of CS in buffering, re-interpreting and transforming problems (Niemiec, 2020), the effect of the levels of "wisdom" in reducing SA has not been explored so far. Moreover, the effect of wisdom on other related dispositional predictors of SA (i.e., interest and cognitive competence towards statistics) has not been tested yet. Hence, this study aims to analyze the effect of the CS of wisdom on SA in a sample of Health and Social Science students, as well as to test whether the levels of individual's interest in statistics and cognitive competence (i.e., self-beliefs and perceptions about knowledge and skills to learn statistics) are mediators of this relationship. The main hypothesis is that CS of wisdom would increase the interest towards statistics, predicting a greater cognitive competence, and leading in turn to less SA. Considering the interaction of these set of variables will disentangle whether the CS of wisdom in college students constitutes a psychological resource that buffers from the SA through increasing interest and cognitive competence. If this hypothesis is supported, it will mean that additional educational tools should be used to compensate for the lack of the wisdom in some students in order to increase their interest and cognitive competence.

## *2. Method*

### *2.1. Participants and Procedure*

Our final sample consisted of $194$ social and health science students ($56.7\%$ females), most of them undergraduate ($99.1\%$), from the University of Valencia (UV, $79.4\%$). Their age ranged from $18$ to $42$ ($M$ = $20.87$, $SD$ = $2.71$). Regarding the average marks in statistics, $28.4\%$ reported that their average was "pass" (5-6.99), $54.1\%$ was "remarkable" (7-8.99), and $16\%$ was "excellent" ($\geq 9$). All participants were required to be studying a master's degree or degree in social or health sciences and to have at least one statistics course as a part of their academic curriculum (i.e., $53.1\%$ from social sciences and $46.9\%$ from health sciences). Students were contacted through the recruitment service "Lineex". Participants read and signed the informed consent, completed all questionnaires online via LimeSurvey platform, and then received a € 5 compensation in return.

### *2.2. Measures*

**Statistical Anxiety**. The Statistical Anxiety Scale (SAS, Vigil-Colet *et al.*, 2008; Oliver *et al.*, 2014) is a 24-item instrument designed to assess three factors of SA towards: (1) Examination, (2) Asking for Help and (3) Interpretation. It is responded on a 5-point Likert scale (1 = "no anxiety"; 5 = "considerable anxiety"), with higher scores indicating higher levels of SA. For this sample the mean score was *M* = $2.90$ (*SD* = $0.63$) and internal consistency was adequate ($\alpha$ ranging from $.82$ to $.92$).

**Attitudes towards statistics**. The Survey of Attitudes Toward Statistics (SATS-36, Schau (*et al.*, 1995; Schau, 2003; Rodríguez-Santero & Gil-Flores, 2019) is a 36-item measure that includes six factors (Affect, Cognitive Competence, Value, Difficulty, Interest, Effort). Items are answered in a 7-point Likert scale (1="completely disagree"; 7 ="completely agree"). For this sample the mean score in Interest was *M* = $5.11$ (*SD* = $0.96$) and in Cognitive Competence was *M* = $5.00$ (*SD* = $1.31$). Internal consistency was adequate ($\alpha$ ranging from $.78$ to $.88$), except for "Difficulty" ($\alpha$ = $.39$).

**Character Strengths**. The Values in Action Inventory of Strengths (VIA-IS; Peterson *et al.*, 2005; Azañedo *et al.*, 2014), is a self-report questionnaire where respondents report to what extent the statements apply to themselves. We assessed only 5 subscales (10 items each) of the questionnaire to assess the Virtue of Wisdom and Knowledge: Creativity, Curiosity, Love of learning, Perspective, and Judgement. This questionnaire uses a 5-point Likert scale (1= "very much unlike me"; 5 = "very much like me"), with higher scores indicating a greater wisdom. For this sample, the mean score was *M* = $37.67$ (*SD* = $3.96$) and the internal consistency was adequate ($\alpha$ ranging from $.72$ to $.90$).

## *3. Results*

A serial multiple mediator analyses were performed with the macro PROCESS version 3.5 for SPSS v.26 (Hayes, 2017). Model 6 was chosen to test whether the association between virtue of wisdom (VIA-IS) and SA (SAS) was mediated by interest and cognitive competence (SATS-36). The mediational effect was significant when the $95\%$ confidence intervals (CIs) of the indirect effect did not include the zero-value. Percentile bootstrap $95\%$ CI based on $5,000$ samples were used to assess the indirect effects.

The indirect effects of model 1 (the single mediation effect through interest), and model 2 (the single mediation effect through cognitive competence) were both not significant. Only the indirect effect of model 3 (the serial mediation effect through interest and cognitive competence) was significant, $\beta = -0.01$, *SE* = 0.00, 95% CI $[-0.02, -0.01]$ (see Figure 1). The multiple regression model including all the variables was significant, $F(3, 190) = 23.45$, $p < .001$), and explained 27.02% of the variance in SA. The direct effect was significant, $\beta = 0.02$, *SE* = 0.01, *p* = .027; while the total effect was not significant, $\beta = 0.01$, *SE* = 0.01, *p* = .336, explaining only 4.8% of the variance in SA.

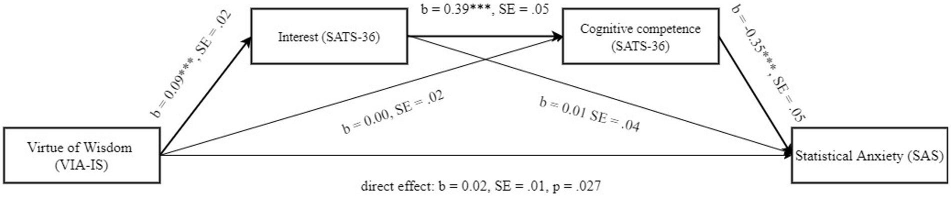


*Figure 1. Serial multiple mediation analyses.*

## *4. Conclusions*

This study was aimed at analyzing the predictor role of the CS of wisdom in SA through its effect on interest and cognitive competence in statistics. Our hypothesis was supported, as the wisdom -the CS that helps people in acquiring and using knowledge- was associated to higher levels of individual's interest in statistics, which was related to higher levels of cognitive competence in statistics, that led in turn to lower levels of SA. Thus, the key role of interest and cognitive competence was in line with previous studies (e.g., Slootmaeckers *et al.*, 2014).

These results have several implications. On the one hand, the CS of wisdom would be acting as a protective factor against SA in individuals that significantly hold this virtue. That is, higher levels of wisdom trigger interest and cognitive competence toward statistics among students that leads to reduce SA. Nevertheless, given the correlational nature of the study, causal predictions can not be inferred. On the other hand, our results point out not only the need of supporting individuals whose CS of wisdom are not as high, but also the direction that these interventions should take in order to reduce SA. Interactive adaptive technological tools, with highly personalised learning path contents, would help to increase interest while making it easier to acquire statistics knowledge and skills. Thus, the development of learning tools as ALEAS app (`https://aleas-project.eu/`) would be effective in targeting core mechanisms in the process of SA, boosting interest and cognitive competence, especially in those students that show lower levels of CS of wisdom. In sum, this is the first study that identifies the CS of wisdom as a dispositional variable that buffers against anxiety towards statistics, highlighting the key role of enhancing interest and cognitive performance in statistics.

*Acknowledgements:* Authors would like to thank the project Erasmus K+ "ALEAS" (European Union, KA2 - Cooperation for Innovation and the Exchange of Good Practices, KA203 - Strategic Partnerships for higher education).

## *References*

Azañedo, C.M., Fernández-Abascal, E.G., Barraca, J. (2014). Character strengths in Spain: Validation of the Values in Action Inventory of Strengths (VIA-IS) in a Spanish sample. *Clinica y Salud*, 25(2), 123-130.

Cui, S., Zhang, J., Guan, D., Zhao, X., Si, J. (2019). Antecedents of statistics anxiety: An integrated account. *Personality and Individual Differences*, 144(February), 79-87.

Duckworth, A.L., Quinn, P.D., Seligman, M.E.P. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology*, 4(6), 540-547.

González, A., Rodríguez, Y., Faílde, J.M., Carrera, M.V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences*, 45, 214-221.

Hayes, A.F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hood, M., Creed, P.A., Neumann, D.L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal*, 11(2), 72-85.

Linley, P.A. (2008). *Average to A+: Realising strengths in yourself and others*. Coventry: CAPP Press.

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

Najmi, A., Raza, S.A., Qazi, W. (2018). Does statistics anxiety affect students' performance in higher education? The role of students commitment, self-concept and adaptability. *International Journal of Management in Education*, 12, 95-113.

Niemiec, R.M. (2020). Six Functions of Character Strengths for Thriving at Times of Adversity and Opportunity: a Theoretical Perspective. *Applied Research in Quality of Life*, 15(2), 551-572.

Oliver, A., Sancho, P., Galiana, L., Iranzo, M.A.C. (2014). Nueva evidencia sobre la statistical anxiety scale (SAS). *Anales de Psicologia*, 30(1), 150-156.

Peterson, C., Seligman, M.E.P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC: American Psychological Association.

Peterson, C., Park, N., Seligman, M.E.P. (2005a). *Assessment of character strengths*. In G.P. Koocher, J.C. Norcross, S.S. Hill III (Eds.), Psychologists' desk reference (2nd ed., pp. 93-98). New York: Oxford University Press.

Rodríguez-Santero, J., Gil-Flores, J. (2019). Actitudes hacia la estadística en estudiantes de Ciencias de la Educación. Propiedades psicométicas de la versión española del Survey of Attitudes Toward Statistics (SATS-36). *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativ*a, 25(1), 1-17.

Ruggeri, K., Diaz, C., Kelley, K., Papousek, I., Dempster, M., Hanna, D. (2008). International issues in education. *Psychology Teaching Review*, 14(2), 65-74.

Schau, C. (2003). Survey of Attitudes Toward Statistics (SATS-36). Retrieved from `http://evaluationandstatistics.com/`

Schau, C., Stevens, J., Dauphinee, T., Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement*, 55(5), 868-875.

Slootmaeckers, K., Kerremans, B., Adriaensen, J. (2014). Too afraid to learn: Attitudes towards statistics as a barrier to learning statistics and to acquiring quantitative skills. *Politics*, 34(2), 191-200.

Vigil-Colet, A., Lorenzo-Seva, U., Condon, L. (2008). Development and validation of the statistical anxiety scale. *Psicothema*, 20(1), 174-186.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. British *Journal of Educational Psychology*, 61, 319-328.

# Statistical Anxiety in undergraduate students: PLS-PM approach for exploring the role of situational and dispositional factors

Rosa Fabbricatore*, Anna Parola**, Giuliana Pepicelli**, Francesco Palumbo**

*Abstract:* This study aimed to examine the role of situational and dispositional antecedents in predicting statistical anxiety (SA). A sample of 253 Italian psychology students were involved. The PLS-PM showed that amotivation, procrastination and test anxiety positively affected SA. Instead, math background, self efficacy and positive affect statistical attitude negatively affected SA.

*Keywords:* Statistical Anxiety, PLS-PM, Higher Education

## *1. Introduction*

Statistics is a fundamental course in most graduate programs, also in human and social science disciplines. Nevertheless, students enrolled in the non-mathematics course often feel discomfort and angst when approaching Statistics. Onwuegbuzie *et al.* (1997) define this apprehension as statistical anxiety (SA), namely *"an anxiety that comes to the fore when a student encounters Statistics in any form and at any level"*.

Over the years, research interest for SA has increased mainly due to its effect on student's performance and academic achievement (Keeley *et al.*, 2008). In particular, studies on the antecedents of SA agree on categorizing the factors inducing SA in *environmental*, *situational* and *dispositional*. Environmental (person-related) antecedents refer to socio-demographic characteristics, such as gender and age. Situational antecedents comprise course-related factors, experience with Statistics, and math background. Finally, dis-

*Department of Social Sciences, University of Naples Federico II, rosa.fabbricatore@unina.it

**Department of Political Sciences, University of Naples Federico II, anna.parola@unina.it, pepicelligiuliana@gmail.com, fpalumbo@unina.it

positional antecedents refer to psychological and emotional factors. Among the dispositional antecedents specifically related to Statistics, the literature showed that negative attitudes toward Statistics play a key role. Moreover, Perepiczka and colleagues (2011) highlighted the importance of students' belief in their competence to face learning statistics' challenges, reporting a significant relationship between self-efficacy and SA.

Furthermore, also dispositional antecedents relating to academic life, in general, have been explored in several studies, i.e. academic procrastination (Walsh and Ugumba-Agwunobi, 2002), academic motivation (Lavasani *et al.*, 2014), and test anxiety (Pintrich *et al.*, 1991). Academic procrastination, defined as the students' tendency to delay academic tasks despite negative consequences intentionally, plays a key role in predicting generalized anxiety, such as test anxiety and social anxiety, and specific kinds of anxiety, such as statistics anxiety. Academic motivation also has a relationship with anxiety during the process of statistics learning. Several studies showed that motivation and introjected motivation often predict negative consequences such as anxiety in the educational context and maladaptive coping with failures (Lavasani *et al.*, 2014). In line with previous studies, these researches highlighted a significant correlation between academic motivation and self-efficacy (Husain, 2014). Moreover, in his studies, Cerino (2014) showed a significant correlation between academic motivation and students' tendency to procrastinate academic tasks.

In this field of study, the construct of test anxiety also plays a central role. Test anxiety is a combination of perceived physiological over-arousal, feelings of worry, tension, and somatic symptoms that occur during test situations (Rajiah *et al.*, 2014), and it is associated with affective and physiological arousal aspects of anxiety (Pintrich *et al.*, 1991).

We aimed to investigate the role of situational and dispositional antecedents in predicting SA. The study considers the three dimensions related to SA: Examination Anxiety, Asking for Help, Interpretation Anxiety. Examination Anxiety and Asking for Help Anxiety do not need to be furtherly detailed. Interpretation Anxiety refers to anxiety in interpreting statistical data and understand the statistical formulation. All the hypothesised relationships are summarised in Figure 1.

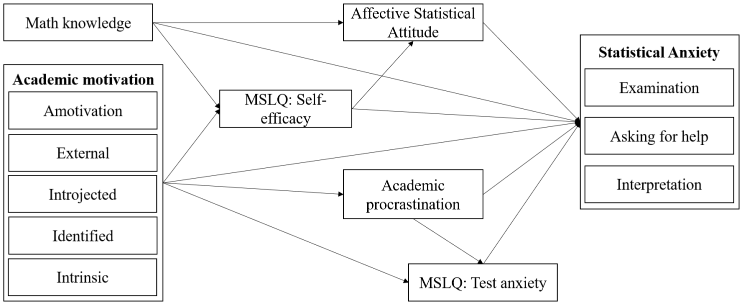


*Figure 1. Hypothesised model*

## 2. *Participants and measures*

Our study involved 253 Italian psychology students attending the first year of university. The following instruments were administered through the Moodle platform: *Statistical Anxiety Scale* (SAS-24; Chiesi *et al.*, 2011), *Mathematical Prerequisites for Psychometrics* (PMP-30; Galli *et al.*, 2008); Affect Statistical Attitude scale (6 items) of *Survey of Attitudes Toward Statistics* (SATS; Chiesi & Primi, 2009); Self efficacy (9 items) and Test anxiety (4 items) scales of *Motivated Strategies for Learning Questionnaire* (MSLQ-44; Bonanomi *et al.*, 2018); *Academic Motivation Scale* (AMS-20; Alivernini & Lucidi 2008); *Tuckman Procrastination Scale* (TP-16; Tuckman, 1991).

## 3. *Statistical analysis*

To test our hypotheses, we exploited a component-based structural equation model (SEM), namely the partial least squares path modelling (PLS-PM; Tenenhaus *et al.*, 2005). SEM theory distinguishes between two conceptually distinct parts of path models: a structural and a measurement model. The former specifies the relationships among the latent variables, and the latter specifies the relationship of the latent to the observed variables. Latent variables can be described and inferred but can not be directly measured; manifest variables are naturally proxies to latent variables and can be directly measured. The PLS-PM iteratively considers the measurement and the structural model during the alternate least squares estimation procedure, providing direct estimates of the latent variable scores as aggregates of the manifest variables. In our work, the relationship between the latent variables and the corresponding indicators was assumed as reflective, as it is typical for psychological constructs. Indicators' reliability, convergent and discriminant validity, and composite reliability were evaluated for the measurement model assessment. The 95% bootstrap confidence interval was used for the structural model to test the regression coefficients' significance. Finally, the $R^2$ statistic was computed for each endogenous variable to measure the percentage of variance explained by the predictors. All the analyses were performed using the R package `plspm` (Sanchez *et al.*, 2017).

## *4. Results and discussion*

Results showed a good quality of the measurement model overall. Indicators' reliability assessment by the factor loadings indicated that almost all items present good reliability (loading $> 0.707$), but one under the minimum threshold of $0.50$. The related item was removed from the procrastination scale in the analysis. The Cronbach's $\alpha$ and Goldstein-Dillon's $\rho$ values were all higher than $0.70$, indicating good internal reliability for all constructs. The measurement model also met convergent validity, as indicated by the average variance extracted (AVE), for which all latent variables presented a value higher or equal to $0.50$. Finally, all the heterotrait-monotrait ratio of correlations (HTMT) values were less than $0.85$, pointing to a good discriminant validity. Since the reliability and validity of the measurement model were ensured, we can consider the structural model. All the significant direct paths among the tested relationships are reported in Figure 2. The $R^2$ values for the endogenous variables ranged from $0.12$ (self-efficacy and test anxiety) to $0.55$ (examination SA), indicating a good proportion of explained variance. Due to the lack of space and considering our goal, in what follows, we focused only on the significant total effects of the considered predictors on SA. About the variables related to the academic life, results showed that amotivated students report a higher level of examination and interpretation SA ($\beta = 0.13$, 95% CI = $0.014$ to $0.26$ and $\beta = 0.15$, 95% CI = $0.04$ to $0.27$, respectively), as well as who internalised an extrinsic academic motivation (AM identified) shows a higher level of examination SA ($\beta = 0.25$, 95% CI = $0.04$ to $0.47$). On the other hand, introjected motivation affects positively the interpretation SA ($\beta = 0.16$, 95% CI = $0.02$ to $0.28$). Moreover, academic procrastination has a positive effect only on the ask for help SA ($\beta = 0.14$, 95% CI = $0.03$ to $0.27$), whereas test anxiety affects all the SA components ($\beta = 0.41$, 95% CI = $0.31$ to $0.50$ for examination; $\beta = 0.39$, 95% CI = $0.27$ to $0.50$ for ask for help; $\beta = 0.25$, 95% CI = $0.14$ to $0.36$ for interpretation). Regarding the course-related predictors of SA, results highlighted that math background negatively affects interpretation SA ($\beta = -0.25$, 95% CI = $-0.36$ to $-0.16$) and that self-efficacy is negatively related to all SA components ($\beta = -0.26$, 95% CI = $-0.36$ to $-0.17$ for examination; $\beta = -0.19$, 95% CI = $-0.29$ to $-0.06$ for ask for help; $\beta = -0.39$, 95% CI = $-0.50$ to $-0.27$ for interpretation). Finally, a positive affective statistical attitude is related to a lower level of examination and interpretation SA ($\beta = -0.41$, 95% CI = $-0.53$ to $-0.31$ and $\beta = -0.36$, 95% CI = $-0.48$ to $-0.23$, respectively).

These findings are in line with those from the literature and suggest considering the different dimensions of SA in the study of its antecedents. Indeed, our results indicate that situational and dispositional antecedents affect SA dimensions differently. In conclusion, we believe that the study of the predictors of SA and the related outcomes can be useful to improve learning and satisfaction in higher education courses.

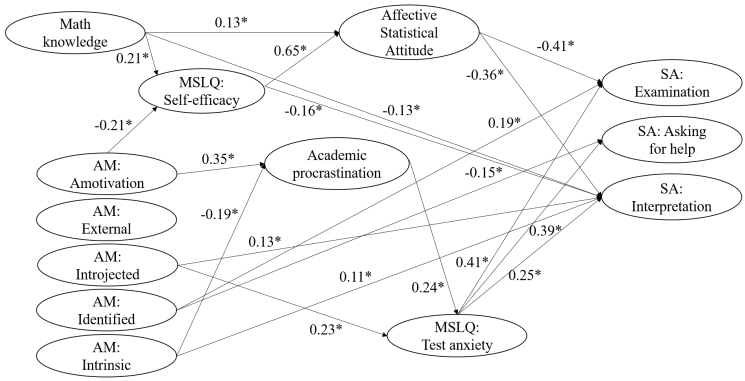


*Figure 2. Structural model: significant direct paths (*p $<0.05$)*

## *References*

Alivernini F., 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.

Bonanomi A., Olivari M.G., Mascheroni E., Gatti E., Confalonieri E. (2018) Using a multidimensional Rasch 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.

Cerino E.S. (2014) Relationships Between Academic Motivation, Self-Efficacy, and Academic Procrastination. *Psi Chi Journal of Psychological Research*, 19(4).

Chiesi F., 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, 309-313.

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

Galli S., Chiesi F., Primi C. (2008) The construction of a scale to measure mathematical ability in psychology students: An application of the Rasch Model. *TPM (Testing Psicometria Metodologia)*, 15, 1-16.

Husain U.K. (2014) Relationship between self-efficacy and academic motivation. *International Conference on Economics, Education and Humanities*, 1(1).

Keeley J., Zayac R., Correia C. (2008) Curvilinear relationships between statistics anxiety and performance among undergraduate students: Evidence for optimal anxiety. *Statistics Education Research Journal*, 7, 4-15.

Lavasani M.G., Weisani M., Shariati F. (2014) The role of achievement goals, academic motivation in statistics anxiety: Testing a causal model. *Procedia-Social and Behavioral Sciences*, 114, 933-938.

Onwuegbuzie A.J., Da Ros D., Ryan J.M. (1997) The Components of Statistics Anxiety: A Phenomenological Study. *Focus on Learning Problems in Mathematics*, 19, 11–35.

Perepiczka M., Chandler N., Becerra M. (2011) Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. *The Professional Counselor*, 1, 99-108.

Pintrich P.R., Smith D.A., Garcia T., McKeachie W.J. (1991) *A manual for the use of the Motivated Strategies for Learning Questionnaire*. Ann Arbor, Michigan.

Rajiah K., Coumaravelou S., 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.

Sanchez G., Trinchera L., Russolillo G. (2017) *Partial Least Squares Data Analysis Methods*. R-package version 0.4-9, `https://CRAN.R-project.org/package=plspm`.

Tenenhaus M., Esposito Vinzi V., Chatelin Y.M., Lauro C. (2005) PLS path modeling. *Computational statistics & data analysis*, 48, 159-205.

Tuckman B.W. (1991) The development and concurrent validity of the procrastination scale. *Educational and psychological measurement*, 51(2), 473-480.

Walsh J.J., Ugumba-Agwunobi, G. (2002) Individual differences in statistics anxiety: The roles of perfectionism, procrastination and trait anxiety. *Personality and Individual Differences*, 33(2), 239-251.

# Understanding students' attitudes toward Statistics

Irasianty Frost*

*Abstract:* Students' attitudes toward statistics are an important factor in learning the subject. Furthermore, attitudes have an impact over time. People with positive attitudes on this matter are more likely to use statistics and deepen their statistical knowledge. The purpose of this study is to gain a better understanding of the development of students' attitudes toward statistics. Moreover, we investigate the mechanisms of how attitudes change. We hope that our findings will support statistics teachers in designing their courses.

*Keywords:* Attitudes Toward Statistics, SATS-36, Bayesian Network.

## *1. Introduction*

Results of extensive research in statistics education strongly indicate that students who have negative attitudes are less willing to put in the effort required for learning statistics. Thus, they prefer a more superficial learning approach (Chiesi & Primi, 2018). Negative attitudes are among the most frequent learning obstacles (Vanhoof *et al.*, 2011). Conversely, students with positive attitudes toward statistics usually achieve better results in this subject (Emmioğlu & Capa-Aydin, 2012). These students also have more self-confidence, less anxiety, and show more interest and more engagement in their learning behaviors. In addition, attitudes have an impact over time (Ramirez, Schau, & Emmioğlu, 2012).

A person's attitudes toward statistics reflect her or his positive or negative disposition with respect to objects, situations or persons connected with learning statistics. In general, attitudes toward statistics are regarded as a multi-dimensional construct. Especially the approach of Candace Schau and her team is widely used (Gal, Ginsburg, & Schau, 1997; Schau, 2003). According to this approach attitudes toward statistics are determined by the components *Affect, Cognitive Competence, Difficulty, Value, Interest* and *Effort*.

*Fresenius University of Applied Sciences, irasianty.frost@hs-fresenius.de

*Affect* reflects positive and negative feelings of students concerning statistics (in which statistics anxiety is subordinated). *Cognitive Competence* captures students' attitudes toward their own cognitive prerequisites for understanding statistics, *Difficulty* captures how students perceive the difficulty of statistics as a subject, *Value* reflects the valence attached to statistics, *Interest* reflects students' interest in the subject, and *Effort* defines students' required effort for learning statistics. If we set positive attitudes as a goal of a statistics course, we need to understand how attitudes develop. This study aims to contribute to a better understanding of the developmental flow of students' attitudes toward statistics. Using the Bayesian networks approach, it explores potential causal direction of the relationships between the components to attitudes. As Xu and Schau (2019) noted, this area has not yet been adequately explored.

## *2. Participants*

Undergraduate students of economics and business psychology at Fresenius University of Applied Sciences in Berlin, Düsseldorf, Hamburg, Cologne and Munich took part in the study. A total of 269 students took part at the beginning of the semester. Of these, 111 (41.26%) were male, 156 (58%) were female, and 2 individuals (0.74%) indicated diverse. The age ranged from 17 to 30 years old (mean = 20.6; sd = 1.13). Toward the end of each semester, there were a total of 161 participants. These included 98 (61%) women and 63 (39%) men between the ages of 18 and 40 (mean = 21.04; sd = 3.03).

## *3. Measures*

Student attitudes were measured using the 36-item SATS-36 (*Survey of Attitudes Toward Statistics*) scale designed by Schau (2003) and her colleages (`www.evaluationandstatistics.com`). SATS-36 exists as pre and post version, thus it is suitable for measuring changes of attitude. The items are given in a Likert scale ranging from 1 (strongly disagree) to 4 (neither) to 7 (strongly agree). SATS-36 captures the components of attitudes: *Affect* (6 items, e.g. "I am scared by statistics"), *Cognitive Competence* (6 items, e.g. "I can learn statistics"), *Value* (9 items, e.g. "Statistics is worthless"), *Difficulty* (7 items, e.g. "Statistics is a complicated subject"), *Interest* (4 items, e.g. "I am interested in using statistics"), and *Effort* (4 items, e.g. "I plan to attend every statistics class session"). The items capture the same content in the pre and post versions, only the tenses differ: in the pre-test, these are formulated in the present tense, while the post version is converted to the past tense as needed.

*4. Procedure*

Attitudes data have been collected online in the summer semester 2019 and the winter semester 2019/20. In the summer of 2019 the surveys took place at the beginning of the semester in March and towards the end of the semester in June 2019 and in the following winter semester in September and mid-December 2019, respectively.

*5. Data Analysis*

We used Bayesian Networks as a probabilistic model selection method. A Bayesian network (BN) is a graphical representation of a joint distribution visualizing multivariate relationships among the random variables (Koller & Friedman, 2009; Pearl, 1988). The graph used in a BN is a directed acyclic graph (DAG) whose nodes correspond with the random variables in the probability distribution. The pattern of the arcs encodes conditional independence statements about the variables. In particular, the decomposition of the joint distribution satisfies the Markov condition, i.e.

$$P(X_1, X_2, \ldots, X_m) = \prod_{i=1}^{m} P(X_i \mid \Pi_{X_i}),\ \Pi_{X_i} = \{\text{parents of } X_i\}.$$

We say that $X_j$ is a *parent* of $X_i$ if there is an arc from $X_j$ to $X_i$ and $X_i$ is a *child* of $X_j$. Either could be the empty set. The direct influence on each variable $X_i$ thus comes directly from the set of parents of $X_i$.

To learn the network structure from the data we used *tabu search algorithm* with *Akaike Information Criterion* (AIC) as a score function. The algorithm is implemented in the R package `bnlearn` which was developed by Scutari (2010) and is freely available at `http://www.bnlearn.com/`. Based on the resulting network the model parameter will be estimated.

For an appropriate model, the data must be normally distributed or discrete. Because this study cannot presume a multivariate normal distribution of the six components (Mardia's skewness measure is 9.4, $p = 7.3 \cdot 10^{-15}$), the data are categorized for the construction of a BN according to the following scheme: scores between 0.5 to below 3.5 are assigned to the category "very to rather negative", 3.5 to below 4.5 are "neutral" and 4.5 to 7 are "rather to very positive". For *Difficulty*, however, the reverse direction of evaluation applies.

## 6. Results

Figure 1 shows the resulting DAG (for the sake of clarity, the nodes are marked with the initials of the components). We can interpret the appearance of *Affect* as the only leading causal variable in the DAG as an indicator of the crucial role of this factor that needs to be taken into account when designing learning environments. The components *Difficulty* and *Effort* occur as consequential variables. That is, how students perceive the difficulty of statistics depends on their emotional attitude towards statistics. The estimated conditional probabilities are not shown and available from the author.

For the BN of attitudes changes it is useful to limit the data to participants both of the pre- and post-test. Their number amounts to 135. Component-wise changes in attitudes are expressed by differences in component scores (post minus pre-scores). For the design of the BN model, the score differences are first categorized. A score change of at least 0.5 points is considered significant. Thus, a difference $\leq -0.5$ is considered decreasing, $-0.5 < x < 0.5$ is considered stable, and $\geq 0.5$ is considered increasing. Consistent with the structure from the pre-test, the DAG roots in *A_Differ* (changes in *Affect*) and ends in *E_Differ* (changes in *Effort*). This means that changes in *Affect* may result in further changes in the other components. In particular changes in *Affect* lead directly to changes in *CognComp* (students' self-confidence) and to changes in *Diff* (students' perceptions about the difficulty of statistics).

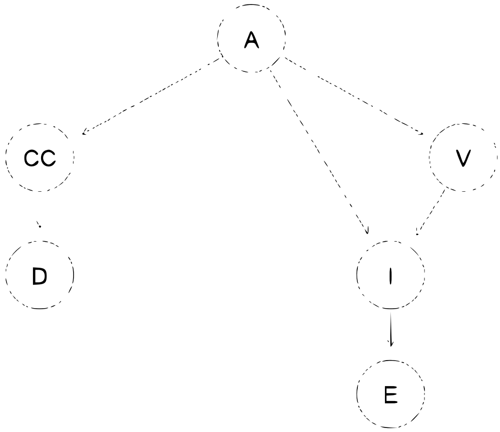


*Figure 1. DAG from the pre-semester data with tabu search and AIC*

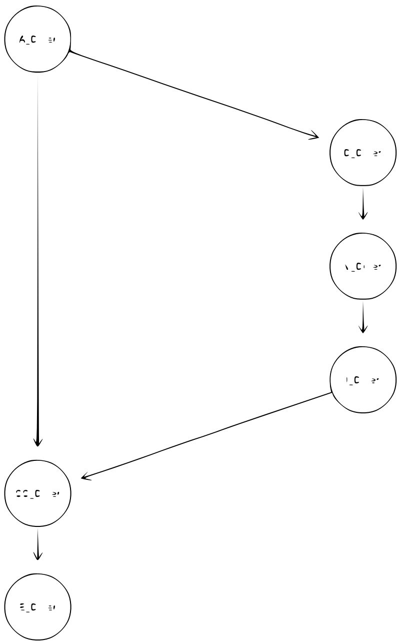

*Figure 2. DAG obtained for the changes in attitudes components*

## *7. Conclusion*

The network structure from the pre-data indicates that *Affect* is crucial in the genesis of students' attitudes toward statistics. This finding is reinforced by the network structure of the score differences of the individual attitude components. We see that changes in *Affect* imply changes in the other components. This outcome should be taken into account when designing intervention strategies. Since *Affect* includes students' emotional attitudes such as joy or fear regarding statistics, it is reasonable to assume that teachers have an influence on it. Further research examining how teachers influence develepemental processes of students' attitudes, development, and evaluation of interventions to improve students' attitudes toward statistics must be pursued.

## *References*

Chiesi F., Primi C. (2018) What happens when attitudes toward statistics change (increase VS descrease) during the course? ICOTS10. `http://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_7D1.pdf?1531364286`

Emmioğlu E., Capa-Aydin Y. (2012) Attitudes and achievment in statistics: A meta-analysis study, *Statistics Education Research Journal*, 11(2), 95-102.

Gal I., Ginsburg L., Schau C. (1997) Monitoring attitudes and beliefs in statistics education. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistics Educations*. IOS Press, S. 37-51. Amsterdam: IOS Press.

Koller D., Friedman N. (2009) *Probabilistic graphical models: Principles and techniques*, Cambridge, Massachusetts, London: MIT Press.

Pearl J. (1988) *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, San Francisco: Morgan Kaufmann.

Ramirez C., Schau C., Emmioğlu E. (2012) The importance of attitudes in statistics education, *Statistics Education Research Journal*, 11(2), 57-71.

Schau C. (2003) Survey of attitudes toward statistics (SAT-36), `http://evaluationandstatistics.com`.

Scutari M. (2010) Learning Bayesian Networks with the bnlearn R Package, *Journal of Statistical Software*, 35(3), 1-22.

Vanhoof S., Kuppens S., Castro Sotos A.E., Verschaffel L., Onghena P. (2011) Measuring statistics attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36), *Statistics Education Research Journal*, 10(1), 35-51.

Xu C., Schau C. (2019) Exploring method effecs in the six-factor structure of the Survey of Attitudes Toward Statistics (SATS-36), *Statistics Education Research Journal*, 18(2), 39-53.

# Section 2
# Tools and methods for the assessment of training paths

# A general nonparametric classification approach for cognitive diagnostic assessment in introductory statistics

Evripidis Themelis*, Angelos Markos*

*Abstract:* Cognitive diagnostic methods (CDMs) are used to assess examinees' mastery and nonmastery of skills to provide formative feedback for teachers and other stakeholders at a finer grain size than a total test score. Nonparametric CDMs were shown to be more efficient than parametric latent class models when either small or large sample sizes are present and more complex sets of skills. In this work, we describe a non-parametric cognitive diagnostic algorithm that improves upon the GNPC (General Non-Parametric Classification) method. The improved GNPC is applied on a real data set to classify a group of undergraduate students based on their answers on a set of descriptive statistics items from the ALEAS App.

*Keywords:* Cognitive diagnostic methods, statistics education, GNPC

## *1. Introduction*

Cognitive diagnosis is a type of assessment typically used in educational measurement to identify an examinee's mastery status of the set of skills required by a set of test problems, based on the examinee's observed scores to these problems. Cognitive diagnostic methods (CDMs) are a class of psychometric tools for cognitive diagnosis modeling (see, e.g. Rupp & Templin, 2008). Instead of overall scores provided in the context of cumulative assessment with the main aim of ranking students based on their total performance, CDMs are tools of formative assessment (Robitzsch & George, 2019). In other words, they provide detailed assessments of the specific strengths and weaknesses in specific skills or attributes, based on which teachers may intervene with each student individually or direct groups of students with a common profile to focused and effective future studies.

CDMs can be built for dichotomous or polytomous data, but in this work we will focus on ordered dichotomous response variables:

*Democritus University of Thrace
ethemeli@eled.duth.gr, amarkos@eled.duth.gr

$$Y_{ij} = \begin{cases} 1, & \text{if subject } i \text{ responds positively to item } j \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1, \ldots, N, j = 1, \ldots, J$. A positive response might be a correct answer on a cognitive test. Instead of a continuous latent unidimensional variable $\theta_i$ expressing an undifferentiated quantity of proficiency or knowledge, CDMs typically contemplate a vector of dichotomous latent variables $\theta_i = (\theta_{i1}, \ldots, \theta_{iK})$ corresponding to $K$ skills or knowledge components needed to respond successfully to test items, where

$$\theta_{ik} = \begin{cases} 1, & \text{if subject } i \text{ possesses skill } k \\ 0, & \text{otherwise.} \end{cases}$$

Thus, subject $i$ is placed into one of $2^K$ latent classes or skill profiles labeled by $\theta_i = (\theta_{i1}, \ldots, \theta_{iK})$, indicating which skills the subject does or does not have.

When applying CDMs to data, researchers (and content experts) posit a hypothesized cognitive structure, known as the $Q$-matrix, indicating the links between the skills of interest and the examinee's observable test responses. A $Q$-matrix is used to map the skill(s) necessary for correct item response onto each item and consists of elements

$$q_{jk} = \begin{cases} 1, & \text{if item } j \text{ requires skill } k \\ 0, & \text{otherwise.} \end{cases}$$

Thus, accurate $Q$-matrix specification is an important aspect.

## 2. *Types of CDMs*

CDMs are derived under assumptions on which skills are needed for which items, and how the skills are utilized to construct a response. *Non-compensatory or conjunctive* CDMs are defined such that all skills associated with an item must be mastered in order for the examinee to have a high probability of answering the item correctly. If the examinee possesses all the skills relevant to an item, the ideal response is $1$ (success); otherwise, it is $0$ (failure). The ideal response profiles are calculated as follows:

$$\eta_{ij}^{c} = \prod_{k=1}^{K} \theta_{ik}^{q_{jk}}. \tag{1}$$

In contrast to the non-compensatory CDMs, *compensatory or disjunctive* CDMs are structured such that mastering a subset of the required skills is sufficient to provide a correct response to the item. This means that only a subset attributes that are measured by an item have to be mastered in order for the examinee to have a high probability of success. In this case, the ideal response profiles are given by:

$$\eta_{ij}^{d} = 1 - \prod_{k=1}^{K} (1 - \theta_{ik})^{q_{jk}}. \tag{2}$$

Another distinction is between *parametric* and *non-parametric* CDMs (Ma, de la Torre, & Xu, 2020). Parametric CDMs are a class of multidimensional, categorical latent-trait models or restricted latent class models, where each class represents a profile of skill mastery (see Rupp & Templin, 2008, for an overview). Given a skill profile for an examinee, the probability of providing a correct response to an item is determined by the skills that are required by the item. Parametric models, however, require a fairly large sample size (often $> 500$ or $1,000$) and their estimation can be time-consuming, making it impractical to apply them to small samples of students, such as those we usually encounter in a school classroom. This gap is filled by *non-parametric* or algorithmic approaches, which, although they do not possess the flexible probabilistic background of parametric models, can efficiently handle small sample sizes and large numbers of skills.

## *3. Nonparametric CDMs for dichotomous responses*

Non-parametric methods avoid the parametric estimation of examinee's proficiency class memberships and instead classify examinees directly into skill profiles based on the calculation of the distance between observed item response vectors and ideal response patterns.

### 3.1. NPC and GNPC

The non-parametric classification (NPC) method (Chiu & Douglas, 2013) assumes either conjunctive, $\eta_{ij}^c$, or disjunctive, $\eta_{ij}^d$, ideal responses and proficiency-class membership is determined by comparing an examinee's observed item response vector $Y_i$ with each of the ideal item response vectors, $H_l = (\eta_{l1}, \ldots, \eta_{lJ})$ of the realizable $2^K$ proficiency classes. An examinee's skill profile $\hat{\theta}_i$ is defined as the skill profile underlying the ideal item response profile, which among all possible ideal item response profiles is closest -or most similar- to the observed item response profile (Chiu & Douglas, 2013):

$$\hat{\theta}_i = \underset{m \in (1,2,\ldots,2^K)}{\arg\min}\ d(Y_i, H_m). \tag{3}$$

A natural choice for $d(Y_i, H_m)$ is the Hamming distance, defined as the number of disagreements between two binary vectors or a weighted Hamming distance that adjusts for different levels of variability in the item responses. Although the NPC assumption is easy to fulfill when the underlying model is purely conjunctive or disjunctive, it is restricted with more complex underlying CDMs (Chiu, Sun, & Bian, 2018).

The general non-parametric classification (GNPC) method (Chiu, Sun, & Bian, 2018) improves upon the NPC by computing a weighted version of the ideal response profiles, as a convex combination of $\eta_{ij}^c$ and $\eta_{ij}^d$:

$$\eta_{ij}^w = w_{ij} \cdot \eta_{ij}^c + (1 - w_{ij}) \cdot \eta_{ij}^d = \eta_{ij}^d + w_{ij} \cdot (\eta_{ij}^c - \eta_{ij}^d).$$

The loss function is then $\underset{m \in (1,2,\ldots,2^K)}{\arg\min}\ d(Y_i, H_m^w)$ and $\hat{w}_{ij} = \dfrac{\sum_{l \in C_l} (y_{ij} - \eta_{lj}^d)}{\|C_l\| \cdot (\eta_{lj}^c - \eta_{lj}^d)}$, where $\|C_l\|$ is the number of examinees belonging to the class of ideal profiles $C_l$. Notice that $\hat{w}_{ij}$ is chosen so that it minimises the Euclidean distance.

### 3.2. Improved GNPC

The GNPC was shown to have similar or greater accuracy than parametric CDMs, especially when sample size was relatively small (Chiu, Sun, & Bian, 2018). However, the choice of the Euclidean distance is not the only option available. A series of simulations have shown that other distance metrics perform significantly better in recovering the true skill profiles (results not shown here for sake of brevity). The best results were obtained for the squared $\chi^2$ distance,

$$d_{Sq\chi^2}(Y, H^w) = \sum_{j=1}^{J} \frac{(Y_j - \hat{H}_j^w)^2}{Y_j + \hat{H}_j^w},$$

which improves the performance of GNPC by $2\%$ to $3\%$, on average.

## *4. Application*

A subset of 18 ALEAS items that require making judgements (Davino *et al.* 2020) were administered to 173 undergraduate students of the Dept. of Primary Education at the Democritus University of Thrace (Greece). The students had recently completed an introductory statistics course. The $Q$-matrix was created by two lecturers in statistics and contained the following five skills (which were also among the course's learning goals): A1: Interpreting graphs for bivariate data (scatterplot, boxplot), A2: Interpreting measures of central tendency and dispersion, A3: Interpreting Correlation, A4: Interpreting Contingency Tables, A5: Normal Distribution (properties and applications).

Students were classified into 32 skill mastery profiles using an R implementation of the revised GNPC. The observed frequencies of the different skill profiles are reported in Table 1. For example, no student has mastered all five skills (profile 11111), whereas 11 participants (6.4%) did not master any of the five skills (profile 00000). Also, about one out of three participants have mastered a single skill only, e.g. 27 (15.6%) A5, 18 (10.4%) A1, 8 (4.6%) A3. With regard to the percentage of mastery for each individual skill, just under one-fourth of the participants (23.7%) have mastered skill A2, about one out of three have mastered skill A4 (34.7%), and less than half of the students have mastered sklls A5 (42.2%), A1 (44.5%) and A3 (45.7%).

*Table 1. Distribution of student profiles across skill mastery levels*

| Skill profile | Number of students | Skill profile | Number of students | Skill profile | Number of students |
|---|---|---|---|---|---|
| 00000 | 11 | 11010 | 1 | 01101 | 3 |
| 10000 | 18 | 00110 | 8 | 11101 | 0 |
| 01000 | 1 | 10110 | 4 | 00011 | 2 |
| 11000 | 6 | 01110 | 9 | 10011 | 10 |
| 00100 | 8 | 11110 | 0 | 01011 | 1 |
| 10100 | 5 | 00001 | 27 | 11011 | 0 |
| 01100 | 6 | 10001 | 2 | 00111 | 3 |
| 11100 | 9 | 01001 | 1 | 10111 | 8 |
| 00010 | 4 | 11001 | 0 | 01111 | 0 |
| 10010 | 6 | 00101 | 8 | 11111 | 0 |
| 01010 | 4 | 10101 | 8 | | |

## 5. Conclusions

In this work, a modified version of the GNPC method allowed for estimating which skills each student has mastered or failed to master in an introductory statistics course. This has provided immediate feedback on students' strengths and weaknesses in the knowledge domain in terms of information learned and information the student still needs to study. Moreover, we can infer the extent to which the instructional goals, as defined by the curriculum, have been accomplished based on which skills the students have collectively mastered. Repeated monitoring of students' skills based on CDMs, can lead instructors to more safe conclusions about whether or not their students possess the skills and knowledge that they has been taught. In this context, the GNPC can be also utilized in computerized adaptive testing due to its computational attractiveness.

## References

Chiu, C.-Y., Douglas, J. A. (2013) A nonparametric approach to cognitive diagnosis by proximity to ideal response profiles, *Journal of Classification*, 30, 225-250.

Chiu, C.-Y., Sun, Y., Bian, Y. (2018) Cognitive diagnosis for small educational programs: The general nonparametric classification method, *Psychometrika*, 83, 355-375.

Davino, C., Fabbricatore, R., Pacella, D., Vistocco, D., Palumbo, F. (2020) ALEAS: a tutoring system for teaching and assessing statistical knowledge, In: Gigliotta, O. and Ponticorvo, M. (eds.) *Proceedings of the Second Symposium on Psychology-Based Technologies*, Naples, Italy, 2730.

Ma, C.,de la Torre, J., Xu, G., (2020) Bridging Parametric and Nonparametric Methods in

Cognitive Diagnosis, arXiv preprint arXiv 2006.15409.

Robitzsch, A., George, A. C. (2019) The R package CDM for diagnostic modeling, In: von Davier, M. and Lee, Y.-S. (eds.), *Handbook of Diagnostic Classification Models*, 549-572.

Rupp, A.A., Templin, J. L. (2008) Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art, *Measurement*, 6(4), 219-262.

# Challenging in teaching biostatistics in an e-learning environment. The experience of a postgraduate course

Claudia Franceschini*, Danila Azzolina**, Ileana Baldi*, Giulia Lorenzoni*, Corrado Lanera*, Paola Berchialla***, Dario Gregori*

*Abstract:* In the last decade, with the spread of new IT (information technology), distance learning diffusion (e-learning) has also increased. However, the online teaching experience is fundamentally distinct from the instruction delivered in a face-to-face setting. The challenge for teaching Biostatistics is harder to health professionals who are low motivated to learn statistics in traditional academic courses. The purpose of this study is to describe the experience of two postgraduate teaching courses in statistics delivered completely online at the Biostatistics Unit (University of Padua). In the 2020-2021 edition, seventy-six students participate to the basic course of Biostatistics, while sixteen applied for the advanced one. Students are overall satisfied with the educational approach; they appreciate the contents, the modules and themes within the courses. However, one of the students' limits is due to a lack of preliminary knowledge about biostatistics notions and an excessive workload in the advanced course. Evaluating student satisfaction regarding e-learning education has taught us that monitoring student satisfaction can be a starting point for redefining the educational setting for future postgraduate editions.

*Keywords:* E-learning, Biostatistics, Postgraduate course.

## *1. Introduction*

In the last ten years, together with the spreading of new IT that facilitates communication through electronic platforms, the demand and diffusion of distance learning (e-learning) have also increased (Yang, 2017). The extensive use of smartphones and mobile applications renders online courses

*Unit of Biostatistics, Epidemiology and Public Health, University of Padua, claudia.franceschini@unipd.it, ileana.baldi@unipd.it, giulia.lorenzoni@unipd.it, corrado.lanera@unipd.it, dario.gregori@unipd.it

**Unit of Biostatistics, Epidemiology and Public Health, University of Padua; Department of Medical Science, University of Ferrara, danila.azzolina@unife.it

***Department of Clinical and Biological Science, University of Turin, paola.berchialla@unito.it

appealing to remote learners, promoting the concept of learning anywhere and everywhere (Song *et al.*, 2004). The greater part of higher education institutions perceives online learning as part of their strategic developing plan.

Online learning courses increased, with the final goal of maximizing the learning experience and reaching more students (Allen & Seaman, 2017). Also, during the last year, the spread of the COVID- 19 pandemic has accelerated this process by disseminating e-learning to all education levels (Kaup *et al.*, 2020). The pandemic has resulted in significant changes in all facets of our livelihoods. Social distancing and movement restrictive policing have significantly overshadowed traditional educative practices (Kaup *et al.*, 2020).

Courses delivered in e-learning mode are typically delivered via Electronic Learning System (ELS). The ELS is a web-based tool aimed at distributing and managing courses over the web platform (Keis *et al.*, 2017). This system involves the implementation of technologies tailored to deliver the learning content facilitating the two-way interaction between teacher and students (Thanji & Vasantha, 2016). The teaching content is typically provided through whiteboards, chat rooms, polls, quizzes, discussion forums, and surveys. The academic institutions' most common platforms are Microsoft Teams, Google G-Meet, Edmodo, and Moodle. In addition to providing support facilities for teaching, these applications also offer video conferencing systems (Barbera & Clarà, 2012). Nevertheless, the online teaching experience is fundamentally distinct from the instruction delivered in a face-to-face setting (Juan *et al.*, 2011). Moreover, the literature argues that the knowledge and skills developed for face-to-face education are not appropriate training for online teaching (Yang, 2017).

The challenge is furtherly hard for teaching Biostatistics to health professionals and medical students, who are in general very little motivated to learn statistics in the traditional academic courses (Zeimet, Kreienbrock, & Doherr, 2015). In most academic courses for physicians and health professionals, students are expected to be skilled in the integration and critical assessment of scientific evidence. The skills related to the epidemiologic study design and data analysis are generally indicated as prerequisites or delivered in introductory courses as epidemiology or biostatistics (Bland, 2004). However, a large number of students in the medical field do not expect to be exposed to mathematical or statistical issues when they start their university education; they are instead concentrated on learning how to examine and treat patients (Freeman *et al.*, 2008; James *et al.*, 2006). On the other hand, the teaching of statistical bases is often performed by teachers with an exclusively mathematical or statistical background who pay more attention to the underlying mathematical principles instead of also focusing on the concepts relevant to health science students (Miles *et al.*, 2010). This teaching approach overshadows the relationship between inferential statistics in medical fields and fails to motivate students to learn statistics (Duffield, Lissemore, & Sandals, 2003).

As mentioned above, biostatistics' online teaching to physicians and health professionals fits into an already challenging educational setting. In addition to delivering statistical and epidemiological skills, the new challenge is to reshape the teaching experience in a different setting from the traditional face-to-face teaching experience. It is even more demanding to teach statistics and other quantitatively oriented courses completely online because these programs usually require more hands-on activities and live demonstrations (Akdemir, 2010).

The purpose of this study is to describe the experience of two postgraduate teaching courses in statistics delivered completely online at the Biostatistics Unit (University of Padua). It is widely documented in the literature that the monitoring of student opinion can be a useful tool for reorienting and modulating the teaching instrument in a new educational online environment that is more flexible by its nature than traditional teaching methods (Yang, 2017). For this reason, we investigate, by using a dedicated questionnaire, the students' satisfaction about the content, techniques, and notions learned during the postgraduate biostatistics course.

## *2. Materials and methods*

The Unit of Biostatistics, Epidemiology and Public Health, at the University of Padua, offers two entirely online postgraduate courses of Biostatistics since the academic year 2016-2017, which present a more and more increasing number of participants during years going from ten-course subscribers in the first edition to ninety-two in the sixth one. The first course, Biostatistics for Clinical Research and Scientific Reporting, is addressed to health professionals and deals with basic biostatistics notions, data analysis, and guidelines for research methodology; the second one, Advanced Biostatistics for clinical research, includes advanced topics in biostatistics and is aimed to statisticians who want to improve their skills in design and analyze clinical trials with innovative methodologies.

Courses are organized into two macro-phases: the first phase, which corresponds to twenty-five and twenty weeks of teaching respectively, is organized in modules concerning topics like missing data, methodology in clinical research, Bayesian design in clinical research, propensity score, basis of a research protocol, etc.; at the end of each module, the student's evaluation is performed through a homework assignment. The students are evaluated on a qualitative ordinal scale including "insufficient", "fair", "good", and "excellent" modalities. The second phase concerns the development of a final project-work and lasts ten weeks of teaching for both courses.

### *2.1. E-learning online platform*

The computerized e-learning portal for education was based on the Moodle platform. Each student has a personal username and password, which allows him to log-in to Moodle's web page's contents without limits in the number of accesses. A course timetable plan is created and shared with the students at the beginning of the courses. The platform allows the educational and logistic management of the course.

1. From an educational perspective, the face-to-face lesson components were translated into digital and online content and tools. For example, lessons were delivered in streaming videos (10-40 minutes) containing the in-person teachers' explanations; the slides shown during the videos, with the e-learning contents, are accessible for students on the platform: Moodle allows students for documents download and upload. The student evaluation for each module is performed through an assessment test composed of multiple-choice questions to be filled on the e-learning platform. Simulation-based tests with unlimited access have also been provided for the students and practical exercises to be car-

ried out within each module, using R and related applications such as Rcmdr and Jamovi. Periodically, videoconferences between students and teachers are scheduled on the platform to provide rapid assistance to learners offering a dedicated space for student-teacher interaction. Zoom Meeting or Google Meet usually provides videoconferences. On top of the courses' pages, two important tools are available: the classified section, which is fundamental to last-minute communications, and the forum section, where students can discuss exercises or topics covered by teachers.

2. From the logistic perspective, the delivery homework time was planned and made available on the platform. Moreover, the platform's security and quality were periodically tested by monitoring student's access to the platform.

For both courses, at the end of each module, a teaching satisfaction questionnaire has been offered to the students. The questions are proposed on a 1-10 Likert scale where 10 indicates the maximum adherence to the question. The questionnaire topics concern 1) the quality of teaching, 2) the appropriateness of the didactical material, 3) the adequacy of the amount of workload required by the module. An open question has also been proposed, where students are invited to submit personal suggestions and impressions.

### *2.2. Statistical analysis*

According to the base and advanced biostatistics course, a descriptive table concerning the student's descriptions has been reported. The data have been represented in terms of the median (II and III quartiles) for continuous variable and absolute with relative frequency for categorical data. The mean satisfaction score for the students' responses in the sixth edition for each item, has been reported for both courses.

## *3. Results*

In the sixth edition, seventy-six students participate in the basic course of Biostatistics, while sixteen applied for the advanced one; for both courses, the number of women is slightly lower than the males. The median age is about thirty-six years in the basic course and thirty-four years in the second one; for both courses, women are younger than males, with a major difference in the advanced course (Table 2).

*Table 2. Description of participants' features in Biostatistics for Clinical Research and Scientific Reporting.*

| Variable | Valid cases | Biostatistics for Clinical Research and Scientific Reporting | Advanced Biostatistics for clinical research | Overall |
|---|---|---|---|---|
| | - | (N=16) | (N=76) | (N=92) |
| **Gender: Female** | 92 | 44% (7) | 49% (37) | 48% (44) |
| Male | - | 56% (9) | 51% (39) | 52% (48) |
| **Age** | 75 | 30.25/36.00/39.50 | 31.00/34.00/41.00 | 30.50/35.00/40.50 |

By referring to the satisfaction questionnaire, we observed that, for both courses, students appreciate a lot the contents and the organization of each module. By the way, one of the worst limits declared is represented by a lack of preliminary knowledge about biostatistics concepts (mean score 7.6 for the basic course, 6.1 for the advanced one) and the course workload in the advanced course, with a mean score of 6.5.

In general, the satisfaction score for the advanced course is lower in comparison with the basic one (Figure 1).

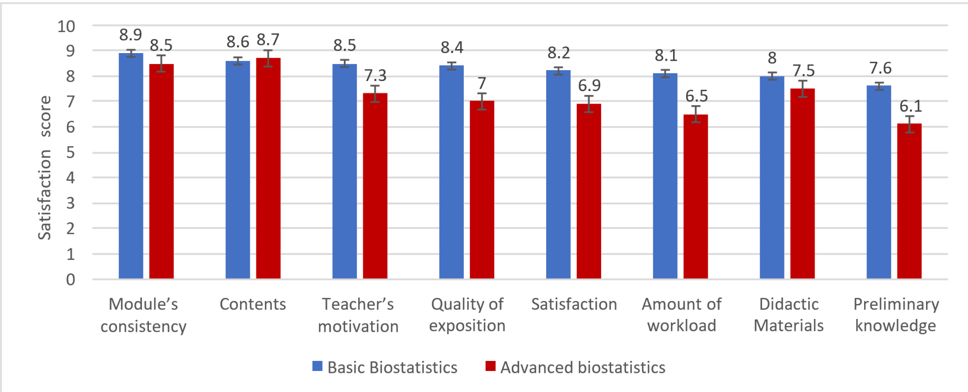


*Figure 1. Average satisfaction score for questionnaire items in Biostatistics courses. Means and error bars have been reported.*

## *4. Discussion*

The experience of postgraduate courses in basic and advanced biostatistics at the University of Padua described in this work served as food for thought about teaching biostatistics to physicians on the e- learning platform. Throughout the various editions, the number of students enrolled has increased, reaching 92 in the sixth edition. This could be due to the general satisfaction of the students about the course organization, but the phenomenon is compliant with the recent diffusion of new flexible e- learning didactic systems, especially for postgraduate courses in medical fields. The e-learning systems in postgraduate medical education have seen a rapid evolution in recent years (Ruiz, Mintzer, & Leipzig, 2006). This type of education offers the students control over the content and learning sequence, allowing them to customize their educational experiences to meet their personal learning goals. The integration of e-learning into medical education may catalyze a shift toward applying an adult learning approach, where educators will no longer serve as content distributors but will be more involved in the process as learning facilitators and skill assessors (Masic, 2008). The recent increase in basic and advanced biostatistics postgraduate course enrollments may also be related to the increased uptake of electronic learning forms during the Covid-19 pandemic. The worldwide spreading of the virus forcibly accelerated, for public health reasons, the learning digitization process at all the educational levels (Amir *et al.*, 2019).

In general, students are satisfied with the organizational aspects of the course, including content. It is very likely that organizing a timeline of work at the beginning of the course, in agreement with the teacher, influences the overall student satisfaction. This organization makes the student responsible for the educational process by giving him/her the possibility to tailor the training time according to his/her own needs (work, family, etc.). The courses also include periodic videoconferences with teachers. This ensures a proactive interaction between the parties involved in the t raining process and is likely to impact the overall student satisfaction levels. Other studies demonstrated that the students conveyed a desire for more in-person interprofessional activities to be included in their education. The Video conferencing systems can facilitate, even if virtually, this process (Palumbo & Bennett, 2016). Regarding biostatistics, most students are not well satisfied with the prior knowledge on the topic. The literature shows that the clinicians often report low confidence and negative attitudes toward statistics. Despite the fact, undergraduate medical schools and postgraduate training curricula involve statistics courses, but, still, many students feel that deficits, stress, and frustration in calculations; affect their competence in statistics (West & Ficalora, 2007).

## *5. Conclusion*

Monitoring student satisfaction can be a starting point for redefining the educational and organizational content of subsequent editions. For example, the course workload load is deemed excessive for the advanced biostatistics course. The content can be redefined whatever necessary for future editions. Also, initial level-up courses could be provided to attempt to compensate for fundamental gaps that hinder learning of statistical topics.

## *References*

Akdemir O. (2010) Teaching Math Online: Current Practices in Turkey, *Journal of Educational Technology Systems*, 39(1), 47-64.

Allen I.E., Seaman J. (2017) Digital learning compass: Distance education enrollment report 2017, *Babson Survey Research Group, e-Literate, and WCET*.

Amir A., Kyriazi P., Lee S.C., Headley D.B., Paré D. (2019) Basolateral Amygdala Neurons Are Activated during Threat Expectation, *Journal of neurophysiology*, 121(5), 1761-1777.

Barbera E., Clarà M. (2012) Time in E-Learning Research: A Qualitative Review of the Empirical Consideration of Time in Research into e-Learning, *ISRN Education 2012*, (january), 1-11.

Bland J.M. (2004) Teaching Statistics to Medical Students Using Problem -Based Learning: The Australian Experience, *BMC Medical Education*, 4(1), 1-5.

Duffield T., Lissemore K., Sandals D. (2003) Teaching the Principles of Health Management to First -Year Veterinary Students, *Journal of Veterinary Medical Education*, 30(1), 64-66.

Freeman J.V., Collier S., Staniforth D., Smith K.J. (2008) Innovations in Curriculum Design: A Multi-Disciplinary Approach to Teaching Statistics to Undergraduate Medical Students, *BMC Medical Education*, 8(1), 1-8.

James E.L., Graham M. L., Snow P.C., Ward B.M. (2006) Teaching Research and Epidemiology to Undergraduate Students in the Health Sciences, *Australian and New Zealand Journal of Public Health*, 30(6), 575-78.

Juan A.A., Steegmann C., Huertas A., Jesus Martinez M., Simosa J. (2011) Teaching Mathematics Online in the European Area of Higher Education: An Instructor's Point of View, *International Journal of Mathematical Education in Science and Technology*, 42(2), 141-53.

Kaup S., Jain R., Shivalli S., Pandey S., Kaup S. (2020) Sustaining Academics during COVID-19 Pandemic: The Role of Online Teaching-Learning, *Indian Journal of Ophthalmology*, 68(6), 1220-21.

Keis O., Grab C., Schneider A., Wolfgang Ö. (2017) Online or Face-to-Face Instruction? A Qualitative Study on the Electrocardiogram Course at the University of Ulm to Examine Why Students Choose a Particular Format, *BMC Medical Education*, 17(1), 1-8.

Masic I. (2008) E-Learning as New Method of Medical Education, *Acta Informatica Medica: AIM: Journal of the Society for Medical Informatics of Bosnia & Herzegovina: Casopis Drustva Za Medicinsku Informatiku BiH* 16(2), 102-17.

Miles S., Price G.M., Swift L., Shepstone L., Leinster S.J. (2010) Statistics Teaching in Medical School: Opinions of Practising Doctors, *BMC Medical Education*, 10(1), 1-8.

Palumbo M.V., Bennett D.N. (2016) Online Video Conferencing: A Promising Innovation in Interprofessional Education, *Journal of Allied Health*, 45(3), 230-234.

Ruiz J.G., Mintzer M.J., Leipzig R.M. (2006) The Impact of E-Learning in Medical Education, *Academic Medicine*, 81(3), 207-212.

Song L., Singleton E.S., Hill J.R., Koh M.H. (2004) Improving Online Learning: Student Perceptions of Useful and Challenging Characteristics, *The Internet and Higher Education*, 7(1), 59-70.

Thanji M., Vasantha S. (2016) ICT factors influencing consumer adoption of ecommerce offerings for education, *Indian Journal of Science and Technology*, 9(32), 1-6.

West C.P., Ficalora R.D. (2007) Clinician Attitudes Toward Biostatistics, *Mayo Clinic Proceedings*, 82(8), 939-943.

Yang D. (2017) Instructional Strategies and Course Design for Teaching Statistics Online: Perspectives from Online Students, *International Journal of STEM Education*, 4(1), 1-15.

Zeimet R., Kreienbrock L., Doherr M.G. (2015) Teaching Biostatistics and Epidemiology in the Veterinary Curriculum: What Do Our Fellow Lecturers Expect?, *Journal of Veterinary Medical Education*, 42(1), 53-65.

# Content validity assessment of ALEAS statistical reasoning items via think-aloud interviews

Aikaterini Tsaousi*, Angelos Markos*

*Abstract:* Over the past decade statistics educators have been interested in assessing cognitive learning outcomes such as statistical literacy, statistical reasoning and statistical thinking. The need for quality assessment tools which measure the concepts that they are intended to measure is increasing. From this point of view, we examined the content validity of a set of ALEAS items for making judgments using think-aloud interviews. Fifteen draft questions were administered to 18 undergraduate students and their cognitive processes while solving them where recorded. A thorough review of student answers revealed unexpected student thinking, as well as wording flaws or poor answer choices that caused confusion to students. This process led to a set of revised questions that can be confidently administered to students to assess their reasoning and improve our understanding of student learning in introductory statistics courses.

*Keywords:* Think-aloud interviews, statistical reasoning, cognitive assessment.

## *1. Introduction*

The field of statistics education has been going through many changes, one of which is a focus on assessments that can reveal important information related to students' learning and thus improve the instruction provided in introductory statistics courses. In particular, there was a shift from computations and procedures to other key learning outcomes: statistical literacy, thinking, and reasoning. In recent years, much attention has been paid to assessing student learning, examining outcomes of courses, aligning assessment with learning goals, and alternative methods of assessment (Garfield & Ben-Zvi, 2008). Currently, there are several assessments available to measure cognitive outcomes of student learning in introductory statistics courses (see Sabbag, Garfield, & Zieffler, 2018 for an overview).

*Democritus University of Thrace
atsaousi@eled.duth.gr; amarkos@eled.duth.gr

During the development of some of those instruments, response process validity evidence was examined via interviews with students who were asked to articulate their thoughts about the instrument items. In fact, Garfield (2003) administered a set of statistical reasoning items to a group of students in the form of open-ended questions and constructed the list of distractors based on their answers. Allen (2006) administered a statistics concepts instrument to a focus group of students and revised it according to their responses. Ziegler (2014) conducted cognitive interviews with students in order to record their thoughts during the Basic Literacy In Statistics assessment and Sabbag *et al.* (2018) used think-aloud interviews in order to examine content validity of the REAsoning and LIteracy assessment. More recently, statistical testing specialists have become interested in think-aloud interview and cognitive laboratory methods, using them both to supplement empirical evidence gathered about students' test item responses and to generate claims that educational and psychological tests measure specific constructs (Leighton, 2017).

The present study aims to assess content validity of a set of items for making judgements developed in the context of the ALEAS project (Davino *et al.*, 2020). "Making judgements" is one of the Dublin Descriptors (Bologna Follow-Up Group, 2005) and refers to "the ability to gather, evaluate, and present information exercising appropriate judgement". This is closely related to the definition of statistical reasoning, defined as the way people reason with statistical ideas and make sense of statistical information (Garfield & Chance, 2000). According to Garfield (2003), this cognitive process involves making interpretations based on sets of data, representations of data, or statistical summaries of data.

## 2. *Think-aloud interviews*

According to Beatty *et al.* (2007), cognitive interviewing has emerged as one of the more prominent methods for identifying and correcting problems with survey questions. Based on their extensive review, there are two major paradigms of cognitive interviewing methods: (a) think-aloud interviewing and (b) verbal probing. We decided to utilize the first one, as during think-alouds the interviewer intervenes as little as possible while participants verbalize their thought processes. Thus, the chances that the interviewer introduces bias into the data collection are eliminated and the thought processes are pure and present less artificiality (Bolton *et al.*, 1996; Conrad *et al.*, 2000).

The think-aloud interview derives from psychological procedures described by Ericsson and Simon (1980). Think-aloud interviews are one-to- one meetings between a research participant and an investigator. The purpose of the interview is to ask the participant to think aloud, which means that the participants articulate their thought as they problems through a series of tasks. The purpose of having participants think aloud is to provide the investigator with the means to identify the thoughts, the cognitive processes and the strategies that participants experience in response to questions or in the course of problem-solving specific tasks (Leighton, 2017).

In fact, recently Reinhart *et al.* (2019) presented a research strategy based upon the think-aloud principle and showed that they are valuable because they reveal that written items frequently do not measure the kind of reasoning that item developers intend them to measure. They observed that students have used reasoning which is unrelated to statistical reasoning, have gotten questions correct despite incorrect statistical reasoning and have revealed new misconceptions. Thus, we expect that think-aloud interviews will provide us with important insights to student thinking while solving ALEAS judgement items and improve our understanding of student learning in introductory statistics courses.

## *3. Method*

### *3.1. Sample and procedure*

A series of think-aloud interviews were conducted with undergraduate students of the Dept. of Primary Education at the Democritus University of Thrace (Greece). The target population was students who had recently completed an introductory statistics course. A group of 38 students who had been enrolled in an elective course on educational measurement were invited to take part in the study. To encourage participation, the tutor (second author) gave a 10% bonus of extra credit. A total of 18 students agreed to take part in the interviewrs which took place during the first and second week of March 2021. Each session was held online due to COVID-19 pandemic restrictions. In each call, there were three people present. The participant, the interviewer and the notetaker.

The think-aloud interview sessions were structured to include 10 minutes for introduction, instructions and a warm-up question to train the participants to think aloud, 30 minutes for 15 multiple-choice questions, and a five-minute period at the end for the interviewer to ask follow-up questions to clarify the student's reasoning where needed.

In line with Reinhart *et al.* (2019), a coding scheme for each response to every question was used to analyze the interview transcripts. Apart from student answers, the method they used to solve each problem was recorded (such as elimination, wording of answer choices, or incorrect statistical reasoning), whether they appeared to misunderstand the question and whether they performed any calculations. A data spreadsheet including all this information was created, along with notes for any unusual method used by students or any revealing comments made during the think-aloud.

### *3.2. Instrument*

The study instrument contained 15 multiple choice questions developed for assessing statistical reasoning in the context of the ALEAS Erasmus+ KA2 project. These questions covered different topics in the learning areas of descriptive and bivariate statistics and were carefully chosen from an initial pool of 50 items, as the most representative in their topic. Each question had four alternative answers of which one is correct and the others reflect common misconceptions.

## *4. Results*

The analysis of the interview data revealed questions that were not interpreted by the students in the way that the questions were designed to be interpreted. Appropriate modifications were then made to these questions, based on student responses. Two examples are provided below.

### *4.1. Ambiguous wording*

The following question requires students to decide which type of graphical display is more suitable to visualize the relationship between two categorical (ordinal) variables.

> A researcher aims to visualize the relationship between the number of words a child uses (low, medium, high) and their family income (low, medium, high). The most appropriate graphical display would be: (a) a histogram for each category of family income. (b) a boxplot for each category of family income. (c) a bar chart for each category of family income [correct answer]. (d) a scatterplot between the number of words and family income. [Learning unit: Graphs for Categorical Data]

In this case, it is reasonable to expect the examinee to eliminate the boxplot, the scatterplot and the histogram, since they are graphical representations appropriate for continuous data. Surprisingly, 12 out of 18 students indicated the scatterplot (d) as the correct answer, whereas 4 students only chose the correct answer. Also, 2 out of 18 chose the “histogram” as the correct answer which was expected because it is a common misconception that students tend to confuse bar graphs with histograms (delMas *et al.*, 2005).

First, notice that number of words and family income are quantitative in nature but in this scenario, they are measured on ordinal scales. Also notice the wording of the four options: the first three contain the phrase “...for each category of family income” while the last one contains the phrase “...between the number of words and family income”. Both these points turned out to be confusing for students. A couple of choice quotes: “I would choose the scatterplot because it is the only display that shows the relationship between two variables” or “in the first three choices, there is only one variable involved, but the fourth mentions them both”. Last, an expert could argue that the best diagram to depict the aforementioned association would be the mosaic plot, but this plot is usually not included in introductory textbooks.

Based on the student feedback, the question was revised as follows: the two ordinal variables were replaced by two “pure” nominal variables, gender and marital status, and the four options were rephrased as (a) a histogram for each gender category (b) a boxplot for each gender category, (c) a bar chart for each gender category income, and (d) a scatterplot for each gender category.

### *4.2. Ineffective distractors*

The results of a survey conducted on a random sample of 1,000 New Yorkers found a statistically significant positive correlation between number of books read and nearsightedness. Which of the following can we conclude about New Yorkers? (a) Reading books causes an increased risk of being nearsighted. (b) Being nearsighted causes people to read more books. (c) We cannot determine which factor causes the other. [correct answer] (d) We cannot draw any conclusions because we need a larger sample size. [Learning unit: Correlation]

In the question above, partially adapted from Reinhart *et al.* (2019), the examinee is expected to apply the rule that correlation does not imply causation and choose option (c), whereas options (a) and (b) refer to a common misconception (X causes Y or Y causes X, if X and Y are correlated). However, 5 students considered 1,000 to be a small sample for the study to be conclusive: "we cannot generalize this result to all residents of New York, given a sample size of 1,000.", and subsequently indicated (d) as the correct answer. Given that the concept under study is correlation, (d) is not an effective distractor, since the examinee should also demonstrate an understanding of the properties of a random sample. Moreover, all 18 students eliminated option (b) simply because it was considered irrational, not because they were applying statistical reasoning about correlation ("This sentence does not make sense and cannot be true").

Based on this feedback, we replaced number of books read with sleep disturbances and nearsightedness with an eye disease, so that a potential bidirectional relationship exists between the two variables. Options (a) and (b) were revised accordingly. Finally, option (d) was replaced by "Both factors cause one another".

## *5. 5. Conclusion*

Think-aloud interviews gave us great insight into student cognitive processes while solving statistical reasoning items. In several cases, students were using different reasoning than we expected, Questions were revised accordingly to improve item quality and enrich our understanding of student learning in introductory statistics courses.

## *References*


Agus M., Peró-Cebollero M., Guàrdia-Olmos J., Penna M.P. (2013, September) The measurement of statistical reasoning in verbal-numerical and graphical forms: a pilot study. In Journal of Physics: Conference Series (Vol. 459, No. 1, p. 012023). IOP Publishing

Beatty P.C., Willis G.B. (2007) Research synthesis: The practice of cognitive interviewing. *Public opinion quarterly*, 71(2), 287-311.

Ben-Zvi D., Garfield J. (2008) Introducing the emerging discipline of statistics education. *School Science and Mathematics*, 108(8), 355-361.

Bologna Follow-Up Group (2005) Framework for Qualifications of the European Higher Education Area. Kopenhagen, p. 9.

Bolton R.N., Bronkhorst T.M. (1996) Questionnaire pretesting: Computer-assisted coding of concurrent protocols.

Conrad F., Blair J., Tracy E. (1999, November) Verbal reports are data! A theoretical approach to cognitive interviews. In Proceedings of the Federal Committee on Statistical Methodology Research Conference (pp. 11-20). Tuesday B Sessions.

Davino C., Fabbricatore R., Pacella D., Vistocco D., Palumbo F. (2020) ALEAS: a tutoring system for teaching and assessing statistical knowledge. *CEUR Workshop Proceedings*, 2730.

delMas R., Garfield J., Ooms A. (2005) Using Assessment Items to Study Students' Difficulty Reading and Interpreting Graphical Representations of Distributions, in K. Makar, ed. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy. Auckland, New Zealand: University of Auckland. Retrieved from `https://www.causeweb.org/artist/articles/SRTL4_ARTIST.pdf`

Ericsson K.A., Simon H.A. (1980) Verbal reports as data. *Psychological Review*, 87, 215-250.

Garfield J., Chance B. (2000) Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning*, 2, 99-125.

Garfield J.B. (2003) Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.

Lane-Getaz S.J. (2013) Development of a reliable measure of students' inferential reasoning ability. *Statistics Education Research Journal*, 12(1).

Leighton J.P. (2017) *Using think-aloud interviews and cognitive labs in educational research*. Oxford University Press.

Reinhart A., Evans C., Luby A., Orellana J., Meyer M., Wieczorek J., ... Nugent R. (2019) Think-aloud interviews: A tool for exploring student statistical reasoning. arXiv preprint arXiv:1911.00535.

Sabbag A., Garfield J., Zieffler A. (2018) Assessing statistical literacy and statistical reasoning: The REALI instrument. *Statistics Education Research Journal*, 17(2), 141-160.

Ziegler L.A. (2014) Reconceptualizing statistical literacy: Developing an assessment for the modern introductory statistics course. Doctorate dissertation. Retrieved by: `https://conservancy.umn.edu/bitstream/handle/11299/165153/Ziegler_umn_0130E_15130.pdf?sequence=1&isAllowed=y`, on 23/01/2021.

# Statistics in Medicine in DAD: first results on experience of Nursing degree courses at the University of Turin

Ilaria Stura* , Alessandra Alemanni**, Giuseppe Migliaretti***

*Abstract:* In March 2020, due to the outbreak of the coronavirus emergency, all the Italian Universities experimented the distance learning (DAD - Didattica a Distanza). In order to evaluate the effects of the DAD on the academic career, the present paper presents some preliminary results relating to the performance of students attending to the Medical Statistics course of the Nursing degree in three campus of the University of Turin in the Academic Years 2019-2020 and 2020-2021. The effect of DAD on student's performance was evaluated using Logistic regression models and the results are showed in terms of OR adjusted for gender, age and campus. The results show that DAD did not bring particular limitations to the students, highlighting on the contrary evident benefits in terms of organization and management of lessons and exams. Moreover, the level of students' satisfaction at the end of the course increased in DAD.

*Keywords:* Distance learning, Academic success, Medical education.

## *1. Introduction*

In March 2020 ('DPCM-8-marzo-2020-1.pdf', no date), due to the outbreak of the Coronavirus emergency, all the Italian Universities experimented the distance learning (DAD - Didattica A Distanza). One of the main questions in this new organization was if DAD will be sufficient for learning new and specific subjects, as required in Degree courses.

The quality of teaching and learning depend on multiple factors, which cannot easily controlled. Many papers, in the last two years, tried to focus on some of them: students' performance (Ziganshin *et al.*, 2020; Foo *et al.*,

*Department of Sciences of Public Health and Pediatrics, University of Turin,
ilaria.stura@unito.it

**Department of Oncology, University of Turin,
alessandra.alemanni@unito.it

***Department of Clinical and Biological Sciences, University of Turin,
giuseppe.migliaretti@unito.it

2021; Jacques *et al.*, 2021), students' perception (Ferraro *et al.*, 2020; Kaila & Kajasilta, 2020; Ramos-Morcillo *et al.*, 2020), teaching and learning instruments (Aleshkovskiy *et al.*, 2020; Casado-Aranda *et al.*, 2020; Co & Chu, 2020), teachers' preparation and motivation (Halpern, 2021).

Moreover, the students' performance is not easy to define. Foo *et al.* (2021) considered standardized scores given by the tutors of the students; Jacques (Jacques *et al.*, 2021) used the median radar diagrams of the students after an engineering project; Ziganshin *et al.* (2020) focused the attention on clinical training.

In this paper, credits achievement is studied in order to assess the students' performance. The achievement of credits cannot be strictly matched with quality of learning or preparation, but it could be a yardstick of the ability of the students to reach the minimum standards that teachers require. In other words, we would like to study the ability of the students on passing the exam in the new framework.

Indeed, two opposite perceptions are present up to now, at least in Italy: someone emphasize the attention deficit in online lessons and the difficulties of the students in understand new subjects in this new framework. For the supporters of this idea, distance learning is not sufficient and a decline of students' performance will be assessed. However, attention deficit is a fact, but do not forget that recordings, books, lecture notes, exercises and other materials are always available to students.

Other people emphasize instead the ease of cheating during online exams, which will lead to an improvement of students' performance. It is true that students are facilitated in coping definitions or theoretical answers. However, online examinations allow the teacher to randomize the questions and make sure that each student has a different question. Moreover, many subjects, especially the mathematical ones, cannot be passed with an only theoretical knowledge. Therefore, coping in these examinations is not very easy.

Considering these two opposite thesis, we want to understand if there is a true difference in credit achievement between in presence and online lessons. As an explorative study, we considered the last two years of one course, the course "Statistics in Medicine" in the degree course of Nursing at the University of Turin.

### *1.1. University of Turin experience*

The University of Turin gave a large variety of instruments for DAD (Didattica alternativa, 2020). Generally, WebEx is the most used platform for lessons and conferences. This platform allows the teachers to both perform lessons simultaneously and register them. During the session, the teacher can share his/her desktop, use a virtual blackboard, have the students' feedback through audio, video and/or chat. Moreover, each course was provided with a Moodle web page in with share WebEx links, WebEx registrations, slides, additional materials (e.g. video-pills), exercises and also exam tests.

As concerns the course of Statistics considered in this paper, during in presence lessons the teacher showed PowerPoint slides and made some exercises and additional explanation on the blackboard. Online lessons were provided with WebEx sessions using PowerPoint slides, the virtual blackboard and Excel sheets for calculations. In Moodle page, both slides and lesson's recording were provided to students.

As concerns the exams, in both cases (on line or classroom mode) students must be connected via WebEx during all the exam, with video and audio turned on. A disconnection of more than five minutes leads to the cancellation of the exam (but there were no cases of this in our samples). Students received the test via e-mail as Word document, then they could both write on the document or on a white paper sheet and take a picture of it. They must create a PDF document with their exam and upload it on a dedicated block in Moodle. The teacher can then see, annotate and correct the files directly on Moodle.

In order to evaluate the effects of the DAD on the academic career, we are developing a study on academic careers at the Degree Courses of Medicine, Nursing and Dentistry at the University of Turin, comparing student's performance by lessons types (in presence or DAD) and by exams types (in presence or online) are evaluated. For now, only the data of the degree course in Nursing relating three campus of the University of Turin (Aosta, Beinasco, Cuneo) are available.

### *1.2. Objective*

The present paper aims to present some preliminary results relating to 308 students attended to the Medical Statistics course of the Nursing degree in three campus of the University of Turin (Cuneo, Aosta and Beinasco) in the Academic Years 2019-2020 and 2020-2021.

## *2. Materials and methods*

In this explorative study, only the Statistics in Medicine course at the course degree in Nursing at University of Turin (Italy) was considered. In particular, three campus were included: Cuneo (Section A and Section B), Beinasco (Section A) and Aosta. The observation period is Academic Years 2019/2020 (first semester) and 2020/2021 (first semester), for the lessons and January 2020, June 2020, July 2020, September 2020, December 2020, January 2021 and February 2021 for the examination sessions. The students were divided in two groups according to the lessons attendance: 'DAD mode' (GroupD) or 'classroom mode' (GroupF). Exams performance were analyzed by exam mode (online vs classroom mode).

The study is based on $308$ students of which $52\%$ belong to GroupD and took the exams in online mode.

In the groupF, the $15\%$ took the exam in online mode, for this reason, it was possible to analyse the effects of DAD both in general (students who attended lessons and took the exam in DAD) and separately only for the exams (students who have attended the lessons 'in classroom mode' but took the exam online mode) in order to see if a ease of coping is really present. Differences between groupD and groupF were evaluated using chi-square test or Mann-Whitney test.

In order to analyze the effect of DAD on exams performance, Logistic regression models were performed and the results are showed in terms of odds ratios (OR) and relative $95\%$ Confidence Intervals ($95\%$CI) adjusted for gender, age and campus.

Finally, students' satisfaction was also considered. After the end of each course, students must compile an evaluation of it and aggregated data are available to teachers by Edumeter. The average of the three campus were considered in order to compare the statisfaction between the two years.

## *3. Results*

A total of $308$ students was analyzed (GroupD: $160$ ($51.9\%$) and GroupF: $148$ ($48\%$)): $20$ from Aosta, $78$ from Beinasco and $210$ from Cuneo campus.

In our sample, $125$ ($40\%$) took the exam in presence and $183$ ($60\%$) in online mode, among the latter $23$ belonged to GroupF. In the first two exam sessions, the $68.3\%$ of the students of GroupF and $71.2\%$ of the students of GroupD were passed ($p = 0.5881$). The $70\%$ of students who took the exam in classroom mode and $68.9\%$ of students who took the exams in online mode passed the exam within the first two sessions ($p = 0.8379$). Therefore, no significant differences were found in passing the exam between lesson and exam types. However, it is interesting to underline that among the students of GroupF who took the exam in online mode, the percentage of students who passed the exam within the first two exam sessions was $89.2\%$.

Even adjusting the estimates for gender, age and campus using Logistic regression models, no statistically significant effects were found between taking exams in classroom mode compared to taking exams in online mode (OR=$0.53$, $95\%$CI from $0.21$ to $1.31$). However, it is worth highlighting the association, also although it is not statistically significant, with passing the exam within the first two exam session with the students of GroupD compared to students of GroupF (OR=$1.79$, $95\%$CI from $0.74$ to $4.38$). The result must also be read in relation to an average improvement in the final evaluations.

Another aspect that is worth highlighting is related to the satisfaction index of students who seem to have statistically increased in online mode, passing from a level of satisfaction average of $80\%$ at a level of $92\%$ ($p = 0.02$).

## *4. Discussion*

The presented results do not highlight significant effects of DAD on student performance. Similar results were reported by (Jacques *et al.*, 2021) and (Ziganshin *et al.*, 2020), while Foo *et al.* (2021) underlined a decrease in performance.

It worth underlining that the investigated subjects are very different. Jacques focused their attention on engineering students; Ziganshin emphasized that the performance of young medical students was unchanged while the older students' one was worsened, as Foo reported considering only medical students at the fourth year.

The difference in these results are therefore due to the subjects: learning mathematical, engineering, basic subjects in DAD is more suitable than learning more specific and/or clinical ones.

For this reason, a more in-depth work is planned by our group in order to compare all the years of the Degree course in Nursing (and maybe in Medicine) and all the subjects. A difference in performance between basic and/or scientific versus clinical subjects is expected.

Another aspect that is worth highlighting is related to the students' satisfaction who seem to have statistically increased levels in comparison to previous years. This is probably due to the greater ease of interaction between teacher and students: indeed, during in presence lessons the students often wait the lesson time to ask questions, or at least they write an email to the Professor. On the contrary, in online mode the teacher can connect via WebEx with a single or a small group of students in order to answer to their doubts, also away from lesson time.

On the other side, some limitations must be underlined. In particular, only the Medical Statistics course was considered, which is not necessarily representative of the whole Nursing degree. However, some positive aspects of the design can be highlighted which have allowed to control the possible bias. For example, the teacher of these courses was the same, and the program, the exam difficulties, books and exercises types were not changed between in presence and in DAD lessons.

Even with all the limitations that our study presents, our study is one of the first aimed at highlighting any effects of DAD on the student's university performance.

The results seem to show that DAD did not bring particular limitations to the students, highlighting on the contrary evident benefits in terms of organization and management of lessons and exams, appreciated by students as evidenced by their level of satisfaction at the end of the course, even in a subject that usually is not particularly popular (especially in clinical curse degree) as the Statistics in Medicine.

*References*


Aleshkovskiy, I.A., Gasparishvili, A.T., Krukhmaleva, O.V., Narbut, N.P., Savina, N.E. (2020) Russian university students about distance learning: Assessments and opportunities, *Vysshee Obrazovanie v Rossii*, 29(10), 86-100.

Casado-Aranda, L.A. Caeiro, S.S., Trindade, J., Paço, A., Casas, D.L., Landeta, A. (2020) Are distance higher education institutions sustainable enough? - A comparison between two distance learning universities, *International Journal of Sustainability in Higher Education*, 22(4), 709-730.

Co, M., Chu, K.M. (2020) Distant surgical teaching during COVID-19 - A pilot study on final year medical students, *Surgical Practice*, 24(3), 105-109.

Didattica alternativa (2020) Università di Torino. Available at: `https://www.unito.it/ateneo/gli-speciali/coronavirus-aggiornamenti-la-comunita-universitaria/didattica-alternativa` (Accessed: 8 March 2021).

'DPCM-8-marzo-2020-1.pdf' (no date). Available at: `https://re.istruzioneer.gov.it/wp-content/uploads/sites/10/2020/04/DPCM-8-marzo-2020-1.pdf` (Accessed: 5 March 2021).

Ferraro, F.V., Ambra, F.I., Aruta, L., Iavarone, M.L. (2020) Distance learning in the covid-19 era: Perceptions in Southern Italy, *Education Sciences*, 10(12), 1-10.

Foo, C.C., Cheung, B. and Chu, K.M. (2021) A comparative study regarding distance learning and the conventional face-to-face approach conducted problem-based learning tutorial during the COVID-19 pandemic, *BMC Medical Education*, 21(1), 1-6.

Halpern, C. (2021) Distant learning: The experiences of brazilian schoolteachers during the covid-19 school closures, *Journal of Ethnic and Cultural Studies*, 8(1), 206-225.

Jacques, S., Ouahabi, A. and Lequeu, T. (2021) Remote knowledge acquisition and assessment during the covid-19 pandemic, *International Journal of Engineering Pedagogy*, 10(6), 120-138.

Kaila, E., Kajasilta, H. (2020) Blended or distance learning?: Comparing student performance between university and open university, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12128 LNCS, 477-484.

Ramos-Morcillo, A.J. Leal-Costa, C., Moral-García, J.E., Ruzafa-Martínez, M. (2020) Experiences of nursing students during the abrupt change from face-to-face to e-learning education during the first month of confinement due to COVID-19 in Spain, *International Journal of Environmental Research and Public Health*, 17(15), 1-15.

Ziganshin, A.M. et al. (2020) Distance learning opportunities during the COVID-19 epidemic, *Kazan Medical Journal*, 101(6), 876-882.

# Differential performance on OECD PISA tests: Evaluating Item Functioning methods using Mathematics Assessments

Clelia Cascella*, Rosa Fabbricatore**

*Abstract:* Differential Item Functioning (DIF) occurs when item parameters change over different sub-groups of subjects. DIF can be manifest and/or latent. Previous studies have shown that DIF can be due to the multidimensionality of the latent trait. In this study, we checked the unidimensionality of the OECD-PISA mathematics achievement test administered in 2018 via the analysis of both manifest and latent DIF. Data analysis showed a few items showing statistically significant and large manifest DIF. Nonetheless, the combined use of both the overall fit statistics and the analysis of the latent DIF confirmed that the unidimensionality of the OECD-PISA achievement test analysed in the current study.

*Keywords:* Rasch model, Unidimensionality, Differential item functioning.

## *1. Introduction: Rasch model in educational assessments*

Rasch modelling is widely used in international large-scale educational assessments such as the Programme for International Students' Assessment (PISA) carried out by the Organization for Economic Cooperation and Development (OECD) (OECD, 2009).

The Rasch model is frequently used because its estimates are invariant, i.e. persons' parameters are test-free and items' parameters are sample-free. Such a property allows for robust comparisons across sub-groups of (i) items, (ii) persons, and (iii) items with persons. In fact, the model postulates that the probability of encountering an item successfully is due to students' relative ability, i.e. their ability compared with the item difficulty, and thus that no other person factors (such as gender, ethnicity, and so on) can affect the probability of encountering an item successfully. In contrast, Differential Item Functioning (DIF; Ackerman, 1992) occurs when groups of subjects with the

*University of Manchester, clelia.cascella@manchester.ac.uk

**University of Naples Federico II, rosa.fabbricatore@unina.it

same ability level have a different probability of correctly answering one or more items. DIF can be manifest and/or latent. The former occurs when groups of respondents are defined according to some manifest variables (such as gender); the latter occurs when there are latent groups of subjects in the population due, for example, to the activation of different cognitive processes.

As stated in Bartolucci *et al.* (2015, p. 205), *"The fact that a given item is affected by uniform DIF may be interpreted in two perfectly equivalent ways. On the one hand, one can think that the difficulty of the biased item is shifted to the right or to the left for one group* [...]. *On the other hand, this is equivalent to assuming two latent traits that are shifted of an amount equal to -c for subjects in group h=2 with respect to those in group h=1. In other words, uniform DIF may be interpreted in terms of multidimensionality, in the case that each ability is perfectly associated with a given group".* Therefore, DIF detection can also be useful to assess the assumption of unidimensionality.

### *1.1. Measuring manifest and latent DIF*

Different approaches have been proposed to check for manifest and latent DIF. The most frequent analysis used to investigate the manifest DIF is the analysis of variance (ANOVA) that reports on the main effect (relative to the stratification variable used to group respondents - 'gender' in our study) and the interaction effect between the factor variable and the student's ability level. The ANOVA is usually combined with the graphical exploration of the item characteristics curves (ICCs) plotted for students belonging to different groups: their comparison allows for the comparison between sub-groups of students matched on the ability and thus identify which groups are disadvantaged compared to the 'reference' group, and for which levels of ability (Cascella *et al.*, 2020).

Regarding the latent DIF detection, the Rasch mixture model (RMM; Rost, 1990) proved to be a suitable statistical tool to detect multidimensionality: if the best model for the data assumes the presence of two or more latent classes for which different Rasch models hold, measurement invariance is violated.

Therefore, we advocate using also RMM as a basic tool for the preliminary assessment of Rasch's assumptions, in addition to the traditional and the most used methods. Indeed, it provides advantages in identifying violations of the measurement invariance, and thus checking for unidimensionality, even when there are no manifest variables for DIF detection.

According to what stated before, in this study, we aimed to check the unidimensionality of the latent trait in OECD PISA mathematics test through the analysis of both manifest and latent DIF. OECD PISA mathematics assessment involves different contents (space and shape, change and relationships, quantity, and uncertainty and data) and various cognitive processes (formulating, employing and interpreting). For this reason, a multidimensional structure could be found.

## 2. Data

OECD PISA assessment aims to measure 15-year-old students' skills and knowledge in reading, mathematics, and science (OECD, 2019). OECD PISA assessment takes place every three years and each round tests one domain in detail. In this study, we focused on the latest round in 2018 when reading was the major domain. OECD PISA uses a rotated booklet design so that students are assigned only a subset of items. In particular, 13 different booklets were administered, each of them representing all the mathematics contents. For our analyses, the 22 items in booklet 2 were considered. Correct answers were coded as 1, whereas wrong answers were coded as 0.

## 3. Statistical methods: Rasch model and DIF detection

In the Item Response Theory (IRT) framework, the Rasch model (Rasch, 1960) defines the probability that the subject $s$ with ability level $\theta_s$ correctly answers the item $i$ as follows:

$$P(Y_{si} = 1|\theta_s, b_i) = \frac{\exp\{\theta_s - b_i\}}{1 + \exp\{\theta_s - b_i\}} \tag{4}$$

where $Y_{si}$ is the response of the subject $s$ to the item $i$ with realization $y_{si} \in [0, 1]$ and $b_i \in \mathbb{R}$ is the item difficulty parameter.

The model introduced by Rasch grounds on three main assumptions: (i) unidimensionality; (ii) monotonicity; and, (iii) local independence. When all these assumptions hold, Rasch estimates are invariant.

A concept clearly related to measurement invariance is that of Differential Item Functioning (DIF): in the Rasch analysis, DIF occurs when item difficulty parameters change over different sub-groups of subjects. Thus, the difficulty parameter $b_i$ in equation 4 is replaced with a class-specific parameter $b_{ik}$, where $k = 1, \ldots, K$ and $K$ is the number of classes.

*Manifest DIF.* In this study, we focused on 'gender' as stratification variable. ConQuest $4.0$ was used to investigate possible differences in difficulty of the items administered to boys and girls. The model statement has three terms that involve two facets, i.e. item and gender. So, as ConQuest passes over the data, it identifies all possible combinations of the item and gender variables. In the current study, it thus constructed $44$ generalised items (i.e., $22$ items by $2$ genders). The model statement requests that ConQuest describes the probability of correct responses to these generalised items using an item main effect, a gender main effect and an interaction between item and gender. The first term yields a set of item difficulty estimates, the second term gives the mean ability of the male and female students and the third term gives an estimate of the difference in the difficulty of the items for the two gender groups.

*Latent DIF.* A suitable statistical tool to detect the latent DIF is the Rasch mixture model (RMM) introduced by Rost (1990). RMM represents a combination of the latent class approach of finite mixture models (McLachlan & Peel, 2000) and the latent trait approach of IRT models, assuming the existence of a different Rasch model for each latent class.

Thus, given $k = 1, \ldots, K$ latent classes, the RMM can be defined as the weighted sum of $k$ Rash models as is shown by the following expression:

$$P(Y_{si} = 1|\theta_s) = \sum_{k=1}^{K} \pi_k \left( \frac{\exp\{\theta_s - b_{ik}\}}{1 + \exp\{\theta_s - b_{ik}\}} \right) \tag{5}$$

where $b_{ik}$ specifies the class-specific difficulty parameter for the item $i$, and the weight $\pi_k$ indicates the prior probability of belonging to the mixture component $k$, with both constraints $\pi_k > 0 \; \forall k \in [1, \ldots, K]$ and $\sum_{k=1}^{K} \pi_k = 1$.

Mixture parameters are estimated by the maximum likelihood via the expectation - maximization (EM) algorithm (Dempster *et al.*, 1977). At the end of the estimation procedure, each subject is assigned to the mixture component for which presents the highest probability of belonging. Regarding the number of latent classes, two selection approaches can be adopted: a priori based on prior knowledge or comparing the fit of models with a different number of latent classes. In the latter case, the BIC can be used as choosing criteria (Cavanaugh, 2016). It is worth noting that latent class DIF modelled by RMM refers to the entire set of test items rather than every single item as is usual in the manifest DIF detection.

## *4. Results and discussion*

Unidimesionality has been investigated first via the mean-square (MNSQ) INFIT and OUTFIT, whose expected value is $1.0$ (Wright & Panchapakesan, 1969), with tolerable standard deviations around $0.20$ (Engelhard, 2009, 2013). In line with the previous studies (Cascella & Pampaka, 2020), we took 1.3 as value for INFIT and OUTFIT mean squares that suggests cause of concern; whereas values below 1 will are not considered as a threat to the measure's validity. Most of the OECD-PISA items show INFIT and OUTFIT mean squares close to 1, with just two exceptions (i.e., the item CM305Q01S and the item CM564Q02S), just slightly above the cut-point of $0.30$. The analysis of manifest DIF shows that female students scored $0.188$ lower than the male students. Such a difference is small (less than $0.20$ logits) but statistically significant. Moreover, even though most of the differences between item parameters estimated for boys and girls are statistically significant, just for one of them (DM446Q02C) DIF magnitude may be considered a cause of concern (Zwick *et al.*, 1999). Latent DIF analysis shows one latent class (no subgroups of students have been found). Indeed, the comparison (based on the BIC) of the fit of models with a different number of latent classes show that the model with one latent class is the best one. Thus, RMM supported the hypothesis of unidimensionality.

## References


Ackerman T.A. (1992) A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective, *Journal of Educational Measurement*, 29, 67-91.

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

Cascella C., Giberti C., Bolondi G. (2020) Studies in Educational Evaluation An analysis of Differential Item Functioning on INVALSI tests, designed to explore gender gap in mathematical tasks. *Studies in Educational Evaluation*, 64, 1-13.

Cascella C., Pampaka M. (2020) Attitudes Towards Gender Roles in Family : A Rasch-based Validation Study. *Journal of Applied Measurement*, 21(2), 1-24.

Engelhard G. (2009) Item and Person Functioning for Students With Disabilities, *Educational and Psychological Measurement*, 69(4), 585-602.

Engelhard G. (2013) *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*, Routledge, London, United Kingdom.

Wright B., Panchapakesan N. (1969) A procedure for sample-free item analysis, *Educational and Psychological Measurement*, 29, 23-48.

Zwick R., Thayer D.T., Lewis C. (1999) An empirical Bayes approach to Mantel-Haenszel DIF analysis, *Journal of Educational Measurement*, 36(1), 1-28.

Cavanaugh J.E. (2016) Model Selection: Bayesian Information Criterion, *Wiley StatsRef: Statistics Reference Online*.

Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society-Series B*, 39, 1-38.

McLachlan G.J., Peel D. (2000) *Finite mixture models*, Wiley-Interscience, New York.

OECD (2009) The Rasch Model in: *PISA Data Analysis Manual: SAS*, Second Edition, OECD Publishing, Paris, `https://doi.org/10.1787/9789264056251-6-en`.

OECD (2019) *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, `https://doi.org/10.1787/b25efab8-en`.

Rasch G. (1960) *Probabilistic models for some intelligence and attainment tests*, University of Chicago Press, Chicago.

Rost J. (1990) Rasch models in latent classes: An integration of two approaches to item Analysis, *Applied Psychological Measurement*, 14, 271-282.

# Section 3
# Technology-based learning experiences

# Investigating the impact of Covid-19 related stress and anxiety on distance learning perception

Maria Iannario*, Alfonso Iodice D'Enza*, Rosaria Romano**

*Abstract:* The purpose of this paper is to understand how high education students perceived distance learning during Covid-19 and whether stress and career-related anxiety impacted their perception. Data are collected from a survey of 1592 students from various Italian Universities and analysed using exploratory factor analysis and partial least squares-based structural equation modelling. The results show a concomitant relationship between student stress and future career anxiety. Besides, anxiety for future career and stress due to fear of contagion do not have a negative impact on distance learning perception of the students. In contrast, a negative relationship is found when considering the effect of stress in relationships and academic life.

*Keywords:* Exploratory Factor Analysis, Structural Equation Modelling, Partial Least Squares.

## *1. Introduction*

The COVID-19 pandemic has affected every aspect of economic and social life, clearly including education. Starting from the second semester of the academic year 2019/20, distance learning (DL) was the only possible way of learning for almost all students of any age. The sudden shift from face-to-face to distance learning undoubtedly affected the learning experience of students. Several social and psychological factors may have influenced students' perceptions of DL.

The aim of this paper is, first, to measure latent concepts such as *student stress*, *future career anxiety*, and *student perspective of DL* using measurement scales already existing in the literature. Second, to extend the discus-

*Department of Political Sciences, University of Naples Federico II,
maria.iannario@unina.it, iodicede@unina.it

**Department of Economics and Statistics, University of Naples Federico II,
rosaroma@unina.it

sion in this field by analysing the relationships between the socio-psychology drivers and the perspective of DL high education students. To achieve this objective, a survey of 1592 students from various Italian Universities conducted in December 2020 is examined. Exploratory Factor Analysis (EFA)(Hair et al, 2006) is used to explore the three scales' latent dimensional structure and Structural Equation Models (SEM)(Bollen, 1989) to investigate the relationships among the identified latent dimensions.

The *Future career anxiety* scale (Mahmud *et al.*, 2020) is conceived to measure a unidimensional conceptualisation of *anxiety* (ANX). It comprises 5 items (Q51-Q55) on a 4-point Likert scale ranging from one ("strongly disagree") to four ("strongly agree"). The *COVID-19 student stress questionnaire* (Zurlo et al, 2020) is assessed to measure the *student stress* (STR) multidimensional conceptualisation. It consists of 7 items on a 5-point Likert scale ranging from zero ("not at all stressful") to four ("extremely stressful"). The seven items are grouped into three subscales: i) four items (Q46-Q49) measure stress concerning relationships with relatives, relationships with colleagues, relationships with professors, and academic studying experience (Relationships and Academic Life - ReAcL); ii) two items (Q45, Q50) measure perceived stress concerning social isolation and changes in sexual life (Isolation - Iso); iii) one single-item (Q44) weighs the stress due to contagion risk (Fear of Contagion - FeCo). The scale used to measure the student's perception of the DL (Amir et al, 2020) consists of twelve items on a 4-point Likert scale ranging from zero ("strongly disagree") to three ("strongly agree"). Here too, the items are grouped into three subscales: 1) two items (Q20, Q21) measure the preference for the DL relative to the clarification sessions and assessments (Preference Domain - PreDom); 2) four items (Q22-Q25) measure the effectiveness of the DL, that is if it creates problems or not, if it causes stress, if it allows you to have more time to prepare learning materials before group discussion or to review all learning materials after class (Effectiveness Domain - EffDom); 3) six items (Q26-Q31) measure satisfaction with the DL (Learning Satisfaction Domain - LsDom).

The remainder of this paper is structured as follows: Section 2 outlines the methodology by including the EFA in Section 2.1 and the SEM in Section 2.2; Section 3 presents the main results and reports some conclusions.

## *2. Method*

### *2.1. Exploratory Factor Analysis*

The three scales' factor structure is assessed through EFA (Spearman, 1904; Hair *et al.*, 2006). It is one of the most widely used statistical techniques in the social and behavioural sciences to measure *constructs* (latent variables, LVs) in surveys. The main idea is that the LV cannot be directly observed, but it has a direct influence on each of the observed indicators (manifest variables, MV) so that they can, in turn, be used to gain insights into the LV.
Given $p$ MVs and $k$ underlying factors, the factor model is:

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{6}$$

where $\mathbf{x}$ is the vector of observed variables, $\boldsymbol{\Lambda}$ is the matrix of regression coefficients (factor loadings) between indicators and factors, $\boldsymbol{\xi}$ is the vector of LVs, and $\boldsymbol{\delta}$ is the vector of uniqueness (unique variances), i.e. variance in the MVs that is not associated with the LVs. The factor model in Eq. 6 can be used to predict the correlation matrix of the MVs as expressed in:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{7}$$

where $\boldsymbol{\Sigma}$ is the model-predicted correlation matrix of the items, $\boldsymbol{\Phi}$ is the correlation matrix for the factors, and $\boldsymbol{\Theta}$ is the diagonal matrix of unique variances. Several extraction methods are available to find estimates of loadings that will yield $\boldsymbol{\Sigma}$ as close as possible to the observed correlation matrix.

### *2.2. Structural Equation Models*

SEM are a class of models for analysing the relationships between LVs that are measured through multiple MVs (Bollen, 1989).

Following the conventional notation, the model can be expressed as

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}, & (8)\\ \mathbf{x} &= \boldsymbol{\Lambda}_{\mathbf{x}}\xi + \boldsymbol{\delta}, & (9)\\ \boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, & (10) \end{aligned}$$

where $\mathbf{y}$ is a $(p \times 1)$-dimensional vector containing $p$ endogenous observed variables, $\mathbf{x}$ is $(q \times 1)$-dimensional vector with $q$ exogenous observed variables, $\boldsymbol{\eta}$ is an $(r \times 1)$-dimensional vector containing $r$ endogenous latent variables, $\boldsymbol{\xi}$ is an $(s \times 1)$-dimensional vector containing $s$ exogenous latent variables; $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are error vectors, respectively, in $(p \times 1)$ dimensions and $(q \times 1)$ dimensions, and $\boldsymbol{\zeta}$ is a residual vector of $(r \times 1)$ dimensions; $\boldsymbol{\Lambda_x}$ and $\boldsymbol{\Lambda_y}$ are respectively loading matrices in $(p \times r)$ and $(q \times s)$ dimensions, and $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are respectively structural coefficient matrices of $(r \times r)$ and $(r \times s)$ dimensions. Both Eqs. 8 and 9 form the *measurement model*, whereas Eq. 10 represents the *structural model*. A LV is defined as *endogenous* if it occurs as a dependent variable in the structural model; otherwise, it is *exogenous*. The same distinction falls on the corresponding MVs, thus distinguishing between endogenous or exogenous MVs.

The estimation methods for SEM follow two different approaches: the covariance-based approach and the component-based approach. The maximum likelihood method is the most well-known estimation methods for the covariance-based approach (Bollen, 1989), whereas partial least squares path modelling (PLS-PM) is the most developed method for the component-based approach (Wold, 1982). The present paper follows the component-based approach since the proposed model is not based on a well-developed and testable theory and a more exploratory approach is thus advisable. The three-step PLSc algorithm (Dijkstra & Henseler, 2015) is used: i) a first iterative phase is carried out to determine the weights to create scores for each construct; ii) the second step corrects for attenuation in correlations between LVs, thus providing a consistent construct correlation matrix; iii) finally, the third step estimates the model parameters (weights, loadings and path coefficients).

## *3. Results*

EFA using *principal axis factoring* extraction method with *oblimin rotation* is carried out on the three scales separately. The Kaiser-Meyer-Olkin measure shows that data are adequate for the factor analysis. It is greater than the minimum acceptable value of $0.5$ for each of the three scales (STR$= 0.76$; ANX$= 0.86$; DL$= 0.93$). *Parallel analysis* is used for determining the optimal number of factors. The results confirm the ANX scale's one-dimensional nature while suggesting to retain two factors for STR and DL instead of the three expected by the original scales. The inspection of the DL *factor loadings* in Figure 1 (left side) shows that the first factor (PA1) corresponds to the LsDom subscale as it is saturated by its respective items but also by two items of the EffDom, which, however, actually concern the sphere of satisfaction (Q22: *I do not experience any problems during DL*; Q23: *I do not experience stress during DL*). The PrefDom subscale is not represented since its indicators (Q20, Q21) are not well explained by the model. The second factor (PA2) reflects the *effectiveness domain*. The EFA results on the STR scale in Figure 1 (right side) suggest removing some items (Q44, Q46, Q49, Q50) since they exhibit weak loadings. Therefore, the first factor corresponds to the ReAcL subscale, while the second to the Iso subscale.

The analysis then continues with the estimation of the model connecting the different subscales using the PLS procedure. The measurement model's assessment confirms *convergent validity*, *composite reliability* and *discriminant validity* (results are available by authors under request). Figure 2 shows the main results of the structural model. The SRMR (standardized root mean square residual) is below the suggested threshold of $0.080$, thus indicating an acceptable model fit. The three dimensions of the STR scale explain $10\%$ of ANX variance, while ANX and STR together explain $32\%$ of EffDom and $46\%$ of LsDom. All direct effects (path coefficients) are significant except the effect of Iso on EffDom. FeCo is the dimension of stress that has the greatest impact on *anxiety*: as stress increases due to the *fear of contagion*, *anxiety for the future career* increases. EffDom depends more on ReAcL: as

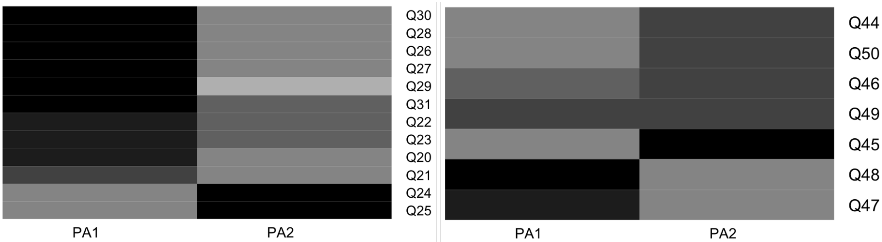


*Figure 1. Factor loadings from the EFA of the DL (left side) and STR (right side) scales, coloured by the intensities of their values*

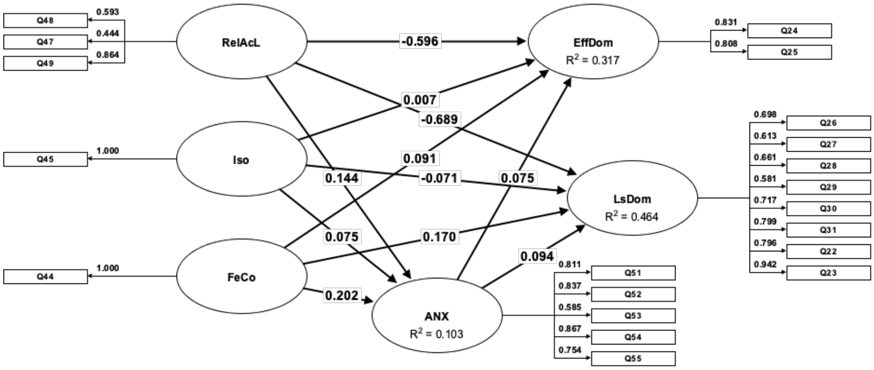


*Figure 2. The results of the full SEM*

stress increases in *social relationships and academic life*, the *effectiveness of DL* decreases. *Learning satisfaction* also decreases with increasing stress in *relationships and academic life* and *social isolation*, while it increases with increasing *fear of contagion* and *anxiety*.

*References*

Amir L.R., Tanti I., Maharani D.A., Wimardhani Y.S., Julia V., Sulijaya B., Puspitawati R. (2020) Student perspective of classroom and distance learning during COVID-19 pandemic in the undergraduate dental study program Universitas Indonesia, *BMC medical education*, 20(1), 1-8.

Bollen K.A. (1989) *Structural Equations with Latent Variables*, Wiley, New York.

Dijkstra T.K., Henseler J. (2015) Consistent and asymptotically normal PLS estimators for linear structural equations, *Computational statistics & data analysis*, 81, 10-23.

Hair J.S., Black W.C., Babin B.J., Anderson R.E., Tatham R.L. (2006) *Multivariate Data Analysis*, Prentice-Hall, New Jersey.

Mahmud M.S., Talukder M.U., Rahman, S.M. (2020) Does 'Fear of COVID-19' trigger future career anxiety? An empirical investigation considering depression from COVID-19 as a mediator, *The International journal of social psychiatry*.

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

Wold H.O.A. (1982) Soft modelling: the basic design and some extensions, in: Jöreskog, K.G., Wold, H.O.A. (eds.), *Systems Under Indirect Observation. Causality, Structure, Prediction*, North-Holland, Amsterdam, New York, Oxford, vol. II, 1-54.

Zurlo M. C., Cattaneo Della Volta M.F., Vallone F. (2020) COVID-19 Student Stress Questionnaire: Development and Validation of a Questionnaire to Evaluate Students' Stressors Related to the Coronavirus Pandemic Lockdown, *Frontiers in Psychology*, 11, 28-92.

# Blended & Data-Driven Learning: a teaching innovation project

Emilia López-Iñesta*, Maria T. Sanz*, Daniel Garcia-Costa*, Francisco Grimaldo*

*Abstract:* The use of technology-based platforms in higher education has increased due to the global pandemic caused by COVID-19, playing a fundamental role in the design of different activities for the monitoring and evaluation of students. This article describes a teaching innovation project framed in the area of Learning Analytics where it is intended to reflect on two aspects: (a) teaching and evaluative adaptation using platforms such as Moodle in blended learning scenarios and (b) the data stored in Moodle platforms that can be used to understand the teaching-learning process. We focus on a teaching experience with primary school teachers in training where the teaching and assessment instruments were adapted due to the COVID-19 scenario. Data analysis suggests that tools that sometimes do not work in pure classroom teaching may play an interesting role in blended learning.

*Keywords:* Moodle, Blended learning, Assessment.

## *1. Introduction*

The educational community had to make a great effort at the beginning of 2020 due to the COVID-19 pandemic to rapidly adapt face-to-face teaching to distance teaching at all educational levels (Bao, 2020). The new circumstances generated a sudden change in the way students were used to receiving classes and evaluating their knowledge. On the other hand, the teaching staff had to modify the programs, changing the contents, the teaching methodology, and the percentages of the continuous and final assessment. This was a preliminary scenario for what is called blended learning, which uses face-to-face and virtual teaching.

In this situation, at the end of the 2019/2020 academic year, a group of teachers from the University of Valencia (UV) in Spain proposed the design

*University of Valencia,
emilia.lopez@uv.es, m.teresa.sanz@uv.es, daniel.garcia@uv.es, francisco.grimaldo@uv.es

of a teaching innovation project framed in the area of data analytics in education or Learning Analytics (hereinafter, LA). The aim is to reflect on two aspects: (a) the adaptation of teaching and the assessment using platforms such as Moodle (López-Iñesta *et al.*, 2015) in blended learning scenarios and (b) how can the data stored in Moodle platforms can be used in understanding the teaching-learning process (Sanz *et al.* 2020; (López-Iñesta *et al.*, 2020 ; Tempelaar, 2020).

Thus, the project "Blended & Data-Driven Learning: Evidence-based on data for blended learning (hereinafter, BDDL)" arose. This paper describes some findings of a teaching experience adapted to the COVID-19 scenario with pre-service primary teachers and preliminary results of Moodle data analysis about the different tools that were used to monitor and evaluate the students in the second half of 2019/2020 university academic year.

## *2. Research questions*

LA can be defined as the area that is responsible for measuring, collecting, analyzing, and presenting data on students and their contexts to understand and optimize learning and the environments in which it occurs (Calvet Liñán *et al.*, 2015; Long *et al.*,, 2011; López-Iñesta *et al.*, 2018, 2020; Romero *et al.*, 2020; Sanz *et al.*, 2020). For this reason, the teaching innovation project described proposes the use of data that are commonly available, but which are usually unknown to teachers. Given the exposed context, two questions are posed to study in this work:

Q1 : How can teaching and assessment be adapted in blended learning settings?

Q2 : Is it possible to use data from student interactions with the teaching platforms used to detect which resources and content work best?

### *2.1. Q1 answer: blended learning and assessment in Higher Education*

Among the most widely used platforms in the field of Higher Education for the creation of "virtual classrooms" where interaction between teachers and students takes place, Moodle, Sakai, Google Classroom, or Chamilo Learning Management System stand out (Bartolomé, 2008). Although its use may seem simple, teachers must carry out correct planning of the blended educational process so that the tasks proposed in the physical classroom or the virtual classroom coexist without inconsistencies between one and the other (Serrate-González *et al.*, 2021). Thus, teachers must not only develop materials and activities for the student to develop independently outside the traditional classroom context but must be trained to take advantage of all the options offered by technological platforms such as making assessments, exchanging files, participating in forums, as well as designing and implementing an important variety of activities.

One of the challenges is monitoring the individual work of students through continuous and final assessment. This poses an even greater challenge for teachers working in classes with a large number of students. Learning Management Systems or teaching platforms such as Moodle play a fundamental role by allowing the design of different activities for the evaluation of students. These include the possibility of creating and importing questionnaires from question banks of different natures: true/false, multiple- choice, matching, short answer, calculation, etc. Problem statements and questions admit tables, mathematical expressions, or images. We design dynamic exercises using the free software R package `exams`. It is possible to create a file in Moodle XML format that can be imported and used in the generation of random questionnaires. In this way, it is possible to have different exercises, but of similar complexity, which facilitates evaluation and makes copying difficult among students. Besides, the completion of individualized questionnaires has led to greater motivation and has encouraged discussion about the problem-solving process among students.

### *2.2. Q2 answer: a preliminary Moodle data analysis study*

Student interactions can help us to know the percentage of access by the students to the Moodle platform system and the dedication to the different activities prepared by the teacher. To facilitate understanding, we look at the different Moodle tools/activities that students have accessed the most during the course in Figure 1.

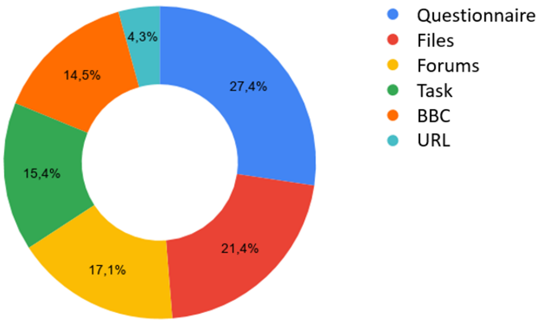


*Figure 1. Student access to the different tools of the platform.*

The access of activities like videoconferences (BBC), tasks, and URLs are in line with the type of course: tasks are mandatory, and students had to deliver it through the platform and BBC sessions, had student's attention that regularly attend face-to-face lessons. However, the participation and access to the forum tool have surprised teachers because it is a tool that on other occasions had no success at all. Throughout the course, four forums corresponding to each of the topics in the subject were proposed. A time- series graph in Figure 2 shows student's access to the different tools used in the course where the highest peaks correspond to the records of the forums tool.

This change can be explained because participation in forums was included in the continuous assessment. This activity score depended on the quality of the answers to the proposed questions and the interaction with the rest of the students in their group.

From this situation, it is possible to deduce in this specific case that a tool such as a forum that is not usually used in person, can reveal interesting details about the participation and interaction between students and teachers. In this sense, Figure 3 represents the meaning of the forum conversations and interactions: the teacher is in the center of the network graph, the students answer the questions proposed by the teacher, but only a few students answer each other (these aspects can be understood from the arrows direction in the network). The size of each node in the network graph is proportional

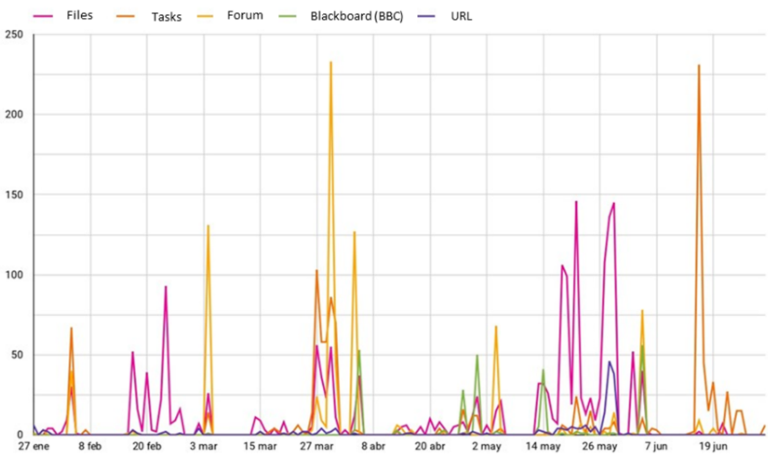


*Figure 2. Records of participation in different course activities.*

to the number of interventions in the forum. Besides, there are some students outside the network who did not participate in the forum activity.


*Figure 3. Word cloud that summarizes the content from the forum activity.*

## 3. Conclusion

In this work, we briefly describe a case study derived from the use of the Moodle platform, which allows the implementation of dynamic exercises and other tools. User interactions with Moodle and the different planned activities generate records whose analysis allows a better understanding of students' study habits and an assessment of teaching practice.

*Acknowledgements:* Authors would like to thank the support of research projects RTI2018-095820-B-I00 (MCIU/AEI/FEDER, UE), EDU2017-84377-R and innovation teaching projects UV-SFPIE _PID19-109833 and UV-SFPIE_PID20-1350001.

## *References*

Bartolomé A. (2008) Entornos de aprendizaje mixto en educación superior. RIED. *Revista iberoamericana de educación a distancia*, 11(1), 15-51.

Bao W. (2020) COVID-19 and online teaching in higher education: A case study of Peking University. *Human Behavior and Emerging Technologies*, 2(2), 113-115.

Calvet Liñán, L., Juan Pérez Á. A. (2015) Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *International Journal of Educational Technology in Higher Education*, 12(3), 98-112.

Long P., Siemens G., Conole G., Gašević D. (2011) Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK11), Banff, AB, Canada, Feb 27-Mar 01, 2011. New York: ACM.

López-Iñesta E., Zacarés M., Grimaldo Arevalillo M. (2015) Generación de eámenes de Estadística para la evaluación continua utilizando R en la plataforma Moodle, *Didáctica de la Estadóistica, Probabilidad y Combinatoria*, 2, 455-456.

López-Iñesta E., García-Costa D., Grimaldo F., Vidal-Abarca E. (2018) Read&Learn: una herramienta de investigación para el aprendizaje asistido por ordenador. *Magister: Revista miscelánea de investigación*, 30(1), 21-28.

López-Iñesta E., García-Costa D., Grimaldo F., Sanz María T., Vila-Francés J., Forte A., Botella C., Rueda S. (2020) Efecto de la retroalimentación orientada al acierto: un caso de estudio de analítica del aprendizaje. *Jornadas sobre Enseñanza Universitaria de la Informática (JENUI)*, 5, 337-340.

Romero C., Ventura S. (2020) Educational data mining and learning analytics: An updated survey. *WIREs: Data Mining and Knowledge Discovery*, 10(3), e1355.

Sanz M.T., López-Iñesta E., Garcia-Costa D., Grimaldo, F. (2020) Measuring Arithmetic Word Problem Complexity through Reading Comprehension and Learning Analytics. *Mathematics*, 8(9), 1556.

Serrate-González S., Torrijos-Fincias P., Gonzále Sánchez M., Caballero Franco D. (2021) Profiling teaching staff using blended learning in their practices in higher education. *Research Papers in Education*, 1-18.

Tempelaar D. (2020) Supporting the less-adaptive student: the role of learning analytics, formative assessment, and blended learning. *Assessment & Evaluation in Higher Education*, 45(4), 579-593.

# "STop Obesity Platform": a gamified learning system to build a new healthier life-style

Monica Zampella*, Eliana Brunetti*, Marianna Mugione*, Raffaele Di Fuccio*, Michela Ponticorvo*, Fabrizio Ferrara*

*Abstract:* STop is a learning system aimed at facing the challenge of obesity through a digital methodology and a synergy between experts from the industry and academia. It will include an innovative platform to collect various People with Obesity (PwO) data from different kind of smart sensor streams and chatbot technology. This data is used as an input for a gamified learning system to build a new healthier lifestyle, through the creation of an app that establishes an analogy to the well-known Picture of Dorian Gray with the purpose of encouraging the performance of the user or the acquisition of healthy behaviour through a stimulating and engaging experience. This paper shows the main objectives of the project, defines the general structure of the app and provides some examples of prototype application.

*Keywords:* Gamification, Learning System, Obesity.

## *1. Introduction*

Obesity is as a complex, multifactorial disorder with both genetic and environmental components, in which abnormal or excess body fat impairs health, increases the risk of long-term medical complications and reduces lifespan. Since 1975, the number of obese people in the world has grown dramatically: Lancet (2016) data reports that obese males have tripled, while obese women have more than doubled. Europe is in second place globally in obesity prevalence levels (Eurostat, 2016; Janssen *et al.*, 2020). This condition does not spare neither children nor adolescents, requiring a multidisciplinary, multi-phase approach (Raj *et al.*, 2010; Güngör, 2014). The interdisciplinary approach in fact is the key to address such a complex issue, characterized by a wide variety of causes and consequences, as well as by a variety of different care pathways.

*Department of Humanities, University of Naples Federico II,
raffaele.difuccio@gmail.com

STop is a project aimed at facing the challenge of obesity through a digital methodology, facilitating the acquisition of healthy habits, replacing unhealthy ones. STop is, in fact, the acronym of “STop Obesity Platform” and consists in an integrate system with an online platform and an app based on gamification, to support and monitor obese people in the path of changing their lifestyle, under the supervision of Healthcare Professionals.

While the platform’s aims are both to collect data on obese patients’ habits and to enable health professionals’ access to them, the app’s objective is to support the increase of self-awareness and self-knowledge in People with Obesity (PwO), leading to the acquisition of healthy habits. For this purpose, the gamification approach was chosen to evoke motivation through a stimulating and engaging experience (Kim, 2012; Cugelman, 2013).

## *2. STop approach to complement obesity treatment*

STOP project is an integrate system based on lifelong learning, which consist of two complementary digital tools. The STOP online platform’s main goal is the acquisition of data on PwO and the tracking of their lifestyle, using different kind of smart sensor streams and chatbot technology. The data will be about four main areas: diet, hydration, walking and exercise; they will be managed through machine learned driven data fusion approaches for sophisticated AI data analysis. Both available data and existing knowledge will allow to personalize goals according to the player’s Body Mass Index (BMI) and the existing habits. All this gathered and analysed data and knowledge will be accessible and usable for Health Care professionals. Furthermore, thanks to the data, the implemented app will offer a personalized user experience, adapted to the real needs of the specific user.

### *2.1. The STop App*

In our work, we design a mobile-app technology based on a gamification approach to facilitate PwO’s adherence to the program and hopefully its long-term maintenance. The app first aim is to promote healthy habits acquisition, focusing on the guidelines found in literature (Nour *et al.*, 2018; Tang *et al.*, 2015).

The second goal of the app is to support people with obesity in the recurrent difficulty of recognizing and managing their emotions (Pink *et al.*, 2019; Baldaro *et al.*, 2003), that may be linked to habits such as overeating (Casagrande *et al.*, 2020; Da Ros *et al.*, 2011). This often creates a vicious circle between confused emotions, emotional eating and weight gain, in which food plays a key role (Fukunishi *et al.*, 1997). Thus, integrated programs involving emotional regulation promotion could be more effective in the long term to improve eating habits, reducing the dropout of participants (Casagrande *et al.*, 2020). Finally, our work also focuses on Nutritional Knowledge, referring to basic information about food components, healthy habits and medical consequences of unhealthy habits. Although the results of many studies on this topic are conflicting, it cannot be excluded that knowledge is one of the elements that contributes to the changing of the habits (Baños *et al.*, 2013), also independently from education and other socioeconomic factors (Bonaccio *et al.*, 2013; Wardle *et al.*, 2000). Basing on this data, we implemented a game section to improve PwO's knowledge on health-related basic information as a tool for promoting conscious healthy choices.

#### *2.1.1 User journey*

The game is intended to last at least 8 weeks, with the possibility to start over again every time the user needs or wants to.

The first aim of the app is to offer a visual representation of the habits' tracking made by the platform. Therefore, in a narrative framework based on the allegory of The Portrait of Dorian Gray, the data about the user's habits in diet, hydration and daily exercise are visually represented in a changing painting; every week a new interactive painting will be unlocked, even if the user doesn't achieve all his/her goals. While the Dorian Gray's portrait shows the direct effect of age and experiences directly on his represented image, in Stop App the effects of the user's effort to start and maintain new healthy habits are showed in a changing natural landscape, leaning on the metaphor of the body as the most important place to live in. This choice was made to help the PwO differentiate their changing process from the body image and the related unrealistic standards promoted by western society.

People with obesity, compared with normal weight individuals, have a higher body dissatisfaction (Cash, 1990; Friedman *et al.*, 2002), such as negative perceptions and feelings about their own body (McGuinness *et al.*, 2016; Slevic *et al.*,, 2011). This is the reason why the Stop App focus on the daily cares their bodies need to be healthy, instead. There are three elements of the painting which constantly change together with the collected data variations: flora, fauna and weather events. The player's results in each one of the four areas will cause a change on the various elements of the painting (i.e., user's eating habits will affect the flora in the picture); the two areas of the exercise and steps are melted together as calories burnt, so they affect just one element of the painting.

During each of the 8 weeks, the players complete a new personalized painting. Furthermore, as the paintings will represent the distinctive specific path of each player, every one of them will have a unique Art Gallery.

After the login, the users will be able to immediately visualize their own art gallery. Through a menu, it will be possible to access the other sections described below.

*Personal area*: it is directly connected with the Stop online platform. The platform collects data about the food eaten, the exercise done, the water drunk, and the steps taken during the day by the user. The player can enter in each area on the platform whenever he/she wants and monitor his/her daily/weekly trend.

*Diary*: it is a section in which the user can talk about emotions, daily achievements and short terms goals. The aim of this area is to make the player think about his/her emotions and express them, thanks to several questions and cues. The data about the emotions could be linked to the nutritional ones, helping the player to find new meaningful connections between them.

*Game*: it is a section based on gamification, designed to cope with the lack of knowledge about healthy habits, so it helps the users to increase their knowledge about healthy and unhealthy habits and their consequences. This activity can also be a fun amusement that will help the user to not give in to his/her moment of weakness.

*Medals section*: also based on gamification, in this section the users can track their successes. The presence of points and/or awards attributed to users' achievements provides immediate feedback on their performance, allowing PwO to focus on even small daily successes and to pay more attention to subsequent goals to maintain high engagement. The users can win up to 4 medals per day, one in each of the four main areas, and occasionally a "Special medal" for different goals: this will help to keep the motivation high, even during rough times.

At the end of the 8th week, user will visualize a week resume to point up both the accomplishments and the areas with more critical issues to work on. It will be also possible to re-start the game and try to achieve more successes.

## 3. Conclusion

STOP platform is an innovative digital system to accompany PwO through the gradual process of improving their health habits. The digital app presented has the aim to increase adherence to treatment and involvement in the use of the platform. To test the usability of the tool and, subsequently, its validity and effectiveness there will be a feasibility study followed by a pilot randomized controlled trial.

In future, the platform may be used in experimental studies to assess its impact on the improvement of health habits, considering variables such as age, gender or social background of participants. The results can then be compared with those obtained through traditional or digital methodologies already in existence.

*Acknowledgements:* Authors would like to thank all the partners of STop Obesity Platform (STOP) Project. This work was supported by the project STOP (STop Obesity Platform) funded by EU under programme Horizon 2020 Research and Innovation - Marie Skłodowska-Curie (Grant number 823978)

## References

Baldaro B., Rossi N., Caterina R., Codispoti M., Balsamo A., Trombini, G. (2003) Deficit in the discrimination of nonverbal emotions in children with obesity and their mothers. *International Journal of Obesity*, 27(2), 191-195.

Casagrande M., Boncompagni I., Forte G., Guarino A., Favieri F. (2020) Emotion and overeating behavior: effects of alexithymia and emotional regulation on overweight and obesity. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 25(5), 1333-1345.

Cash T.F. (1990) The psychology of physical appearance: Aesthetics, attributes, and images. In T. F. Cash, T. Pruzinsky (Eds.), *Body images: Development, deviance, and change*, Guilford Press, 51-79.

Cugelman B. (2013) Gamification: what it is and why it matters to digital health behavior change developers. *JMIR serious games*, 1(1), e3.

Da Ros A., Vinai P., Gentile N., Forza G., Cardetti S. (2011) Evaluation of alexithymia and depression in severe obese patients not affected by eating disorders. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 16(1), 24-29.

Eurostat. European Health Interview Survey (2016) - Almost 1 adult in 6 in the EU is considered obese - Share of obesity increases with age and decreases with education level. Eurostat news release 20 October 2016.

Friedman K.E., Reichmann S.K., Costanzo P.R., Musante G.J. (2002) Body Image Partially Mediates the Relationship between Obesity and Psychological Distress. *Obesity Research*, 10, 33-41.

Fukunishi I., Kaji N. (1997) Externally oriented thinking of obese men and women. *Psychological Reports*, 80, 219-224.

Güngör N.K. (2014) Overweight and obesity in children and adolescents. *Journal of clinical research in pediatric endocrinology*, 6(3), 129-143.

Janssen F., Bardoutsos A., Vidra N. (2020) Obesity Prevalence in the Long-Term Future in 18 European Countries and in the USA. Obes Facts 13, 514-527.

Kim B. (2012) Harnessing the power of game dynamics why, how to, and how not to gamify the library experience. *College & Research Libraries News*, 73(8), 465-469.

McGuinness S., Taylor J.E. (2016) Understanding body image dissatisfaction and disordered eating in midlife adults. *New Zealand Journal of Psychology*, 45(1), 4-12.

Nour M.M., Rouf A.S., Allman-Farinelli M. (2018) Exploring young adult perspectives on the use of gamification and social media in a smartphone platform for improving vegetable intake. *Appetite*, 120, 547-556.

Pink A.E., Lee M., Price M., Williams, C. (2019) A serial mediation model of the relationship between alexithymia and BMI: The role of negative affect, negative urgency and emotional eating. *Appetite*, 133, 270-278.

Raj M., Kumar R.K. (2010) Obesity in children & adolescents. *The Indian journal of medical research*, 132(5), 598-607.

Slevic J., Tiggemann M. (2011) Media exposure, body dissatisfaction, and disordered eating in middle-aged women. *Psychology of Women Quarterly*, 35, 617-627.

Tang J., Abraham C., Stamp E., Greaves C. (2015) How can weight-loss app designers' best engage and support users? A qualitative investigation. *British journal of health psychology*, 20(1), 151-171.

Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants, (2016), 387, 1377-96.

Wardle J., Parmenter K., Waller J. (2000) Nutrition knowledge and food intake. *Appetite*, 34(3), 269-275.

# ALEAS: an interactive application for ubiquitous learning in higher education in statistics

Raffaele Di Fuccio*, Fabrizio Ferrara*, Giovanni Siano**, Andrea Di Ferdinando*

*Abstract:* ALEAS app is a training system that can be used to improve students' work during a typical statistic course. ALEAS supports students' learning providing several resources (tests, videos, cartoons, texts) usable through mobile devices for ubiquitous learning aimed at students in higher education courses in humanities (psychology, education, social sciences, etc.) where statistics is a hurdle to overcame. The app provides an environment where the user is able to train itself, in parallel with the work in the university classroom. In this paper the authors describe the architecture of the application, highlighting the logical blocks and the structure of the application. The app is downloadable, free and open on the Google Play Store.

*Keywords:* blended learning, mobile learning, statistical anxiety.

## *1. Introduction: a mobile application for training in statistics*

ALEAS is a training statistical app designed in order to prevent the anxiety on statistics courses. ALEAS supports students' learning providing several resources (tests, videos, cartoons, texts) usable through mobile devices (Davino *et al.*, 2021). The project ALEAS, funded under the Erasmus+ Programme, aimed at the development of an APP: it is considered crucial to motivate students because mobile devices are indeed the pocketknife of digital natives and important in blended learning (Limone, 2021).

ALEAS is designed with the focus on the application of quantitative methods in real life, stimulating students to appreciate the importance of statistics in their future professions. Such design guideline is intended to face with skills gap and to strengthen the diffusion of statistical literacy (Gould, 2017).

*Smarted srl, raffaele.difuccio@smarted.it, fabrizio.ferrara@smarted.it, andrea.diferdinando@smarted.it

**QUING'S'YSTEMS srl, giovanni.siano@quing.it

The project develops this interactive application for ubiquitous learning for students in higher education courses in humanities (psychology, education, social sciences, etc.) where statistics is a hurdle to overcame. ALEAS exploits videos and cartoons to speak the same language of digital natives. ALEAS Toolbox is a training APP that can be used to improve students' work during a typical statistic course. The APP provides an environment where the user is able to train itself, in parallel with the work in the university classroom.

The design principles of ALEAS provides ad-hoc learning paths so as to offer students a modern mentoring system.

Each interaction of the user foresees the following steps:

1. The educational input, the requests from the system to the learner, an image, a video or aural speech.

2. The user replies to the educational input.

3. The APP elaborates the reply, this module defines the coordination of the outputs and the feedbacks based on the exercise script and the embedded adaptive intelligent tutors.

4. The educational output, it implies the displaying and the production of feedbacks. These could be videos, pictures or voices or a combination of those.

During the interaction between the APP and the student, the system tracks behaviours and analyse results, giving new data in order to present a new exercise based on the "pedagogical history" of the learner and on the setting defined by the professor.

## *2. The APP architecture*

The software architecture consists of a sequence of logical blocks. The main building blocks are depicted in the Figure 1. In order to form a data processing chain, the main building blocks are connected into each other either with a one-to-one relationship or a one-to-many relationship. The descriptions for these blocks and their (initial) icons are given in Table 1.

*Table 3. Description of the graphic elements.*

| Type | Figure | Description |
|---|---|---|
| Data store | Purple cylinder | Contains the (main) data used in the ALEAS project. The contents are statistical exerises provided in the IO2, by the statistical experts and lecturers. Exercises within the data store follows the ALEAS data structure format, done by EXAMS. |
| Filters | Red pentagon | Filters transform the data coming from a data collection, data store or a tool before passing it onward. They are used to validate data. Filters are managed with algorithms that assess which is the level of each students and which is her/his ability to know the new contents. In addition filters are able to provide learning proposals to the students. |
| Learning data | Yellow rectangles | This block provides learning inputs to the students that could improve their competences browsing the slides with the main information of the statistics modules discussed in the APP. |
| Interactive modules | Dotted blue oval | These are the exercises that the user is able to reply using the APP with interactive GUIs. |
| Input | Dotted diamond | The dotted line means that it is needed the input from the student. |

After the user Login, the student will be forced to follow two different paths. If the user is performing his/her first open of the APP proceeds through two blocks. The first one asks demographic data regarding the student, a-priori weaknesses of students are identified according to their features and paste experience (e.g. age, sex, type of diploma, diploma mark). The second one presents a questionnaire in order to assess the anxiety regarding statistics

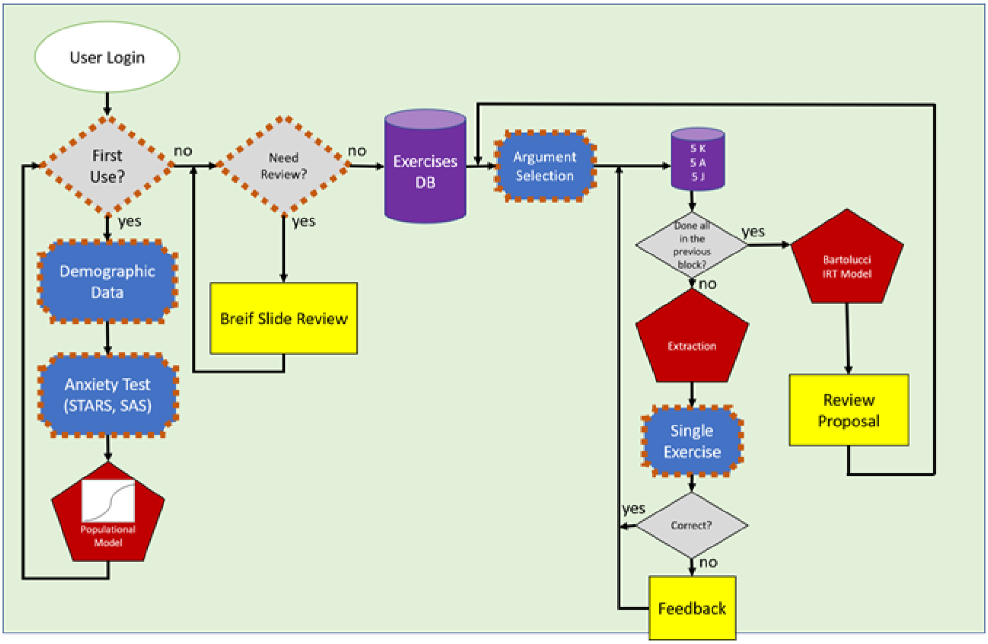


*Figure 1. Architecture of ALEAS*

and mathematic through the SAS (Statistical Anxiety Scale) (Vigil-Colet *et al.*, 2008).

All the data collected feeds an algorithm that is able to classify the student, based on the data collected in the first two activities. In this way these results will be used to define a typology of students and classifying them into homogeneous groups. This a-priori classification is useful to develop a system "tailored" for each group and able to consider heterogeneous statistical skills, attitude and expectations.

Then the student could browse complementary didactic tools for statistics (e.g. slides of the contents where the student could train with the ALEAS APP). This block could be browsed for all the time that the student considers appropriate, or he/she could skip this and proceed with the exercise training.

Next, the student could select the preferred argument in the unblocked exercises. The unblocked arguments are defined on the basis of each student and related to the first login, and the consequential questionnaires, and the replies of the student in the previous exercises. Each time the student reaches the level labelled as "high" in a specific argument; he/she will be endorsed to start with a new argument. However, the choice of the argument is free, and the student has the freedom to continue to train himself/herself in the unblocked group of exercises.

When a block of exercises is selected, it comprises 15 questions, with 5 for each following category (according with the well-known Dublin Descriptors, BWG, 2005), ALEAS is conceived to evaluate and to improve student's knowledge with respect to: i) knowledge and understanding; ii) applying knowledge and understanding; iii) making judgments.

The exercises are divided in: i) item, ii) topic and iii) areas. The item represents the first level of the syllabus and includes from 15 to 45 exercises, uniformly distributed between those where it is necessary to solve an exercises (application), where the students has to reply to theoretical questions (knowledge) and finally, where the user has to provide a choice based on its competences (judgment). Then, a single exercise is extracted from the item and the user is able to reply. If the reply is right, he/she will proceed with the next exercise until the topic is completed; in the case the reply is wrong, the student will be able to have a comment with the explanation of his/her error at the end of the item.

After the item is completed, the system elaborates the Item Response Theory model named Bartolucci model (2017) that will classify the students replies. Then, at the end of each module, based on students' interactions with ALEAS, each student will be assigned to the most likely performance group (e.g. low, medium or high). Her/his position within the students ranking is also identified. The expected result of this phase is a positive difference between the a priori and a-posteriori probability of getting a good performance.

## *3. Conclusion*

The ALEAS project delivered an application for ubiquitous learning using mobile devices. The architecture of the app is designed in order to sustain the students enhancing its motivation with video-cartoons for the explanation of specific arguments, cartoon feedbacks when an item is well performed and using supportive materials in the form of slides and learning materials in order to improve the knowledge in statistics arguments.

The system was tested in the ALEAS project and is downloadable in Google Play store[1] and free to use for every student. In this paper we showed the technical infrastructure, further analysis will be performed for the analysing the learning impact in real higher education environments.

*Acknowledgements:* The ALEAS (Adaptive LEArning in Statistics) [www.aleas-proect.eu] project is funded by the European Commission under the call Erasmus+ KA2 - Strategic Partnership - Higher Education, Grant Agreement n. 2018-1-IT02-KA203-048519. The European Commission's does not constitute an endorsement of the contents of this publication, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

## *References*

Bartolucci F., Bacci S., Gnaldi M. (2017) MultiLCIRT: Multidimensional Latent Class Item Response Theory Models, *R package version 2.11*.

Bologna Working Group (BWG) (2005) A Framework for Qualifications of the European Higher Education Area. Copenhagen: Danish Ministry of Science, Technology and Innovation.

Davino C., Fabbricatore R., Pacella D., Vistocco D., Palumbo F. (2020) ALEAS: a tutoring system for teaching and assessing statistical knowledge, *CEUR Workshop Proceedings*, 2730.

Gould R. (2017) Data literacy is statistical literacy. *Statistics Education Research Journal*, 16(1), 22-25.

Limone P. (2021) Towards a hybrid ecosystem of blended learning within university contexts, *CEUR Workshop Proceedings*, 2817.

Vigil-Colet A., Lorenzo-Seva U., Condon L. (2008) Development and validation of the statistical anxiety scale. *Psicothema*, 20(1), 174-180.

[1] `https://play.google.com/store/apps/details?id=it.smarted.aleas&hl=en_US`

The volume collects a selection of papers presented during the Conference "Stat.Edu'21 - New Perspectives in Statistics Education". The conference was held at the Department of Political Sciences of the University of Naples Federico II (25 - 26 March 2021) as the final event of the "ALEAS - Adaptive LEArning in Statistics", an ERASMUS+ project (`https://aleas-project.eu`) developed in the period $2018 - 2021$.

Applicant Organization: University of Naples Federico II (Italy) - Coordinator: professor Francesco Palumbo
Partners: Jacobs University Bremen, Germany (Adalbert Wilhelm); Democritus University of Thrace, Greece (Angelos Markos); Polytechnic University of Valencia, Spain (Rosa María Baños); Smarted, Italy (Raffaele Di Fuccio).

Università degli Studi di Napoli Federico II
Scuola delle Scienze Umane e Sociali

The volume collects the papers presented at the conference "Stat.Edu'21 -New Perspectives in Statistics Education" which was held at the Department of Political Sciences of the University of Naples Federico II (25-26 March 2021). The conference was the final event of the "ALEAS - Adaptive LEArning in Statistics", an ERASMUS+ project (https://aleas-project.eu) developed in the period 2018-2021 to design and implement an Adaptive LEArning system able to offer personalised learning paths to students, with the purpose to provide them remedial advice to deal with the "statistics anxiety".
Stat.Edu'21 aimed at stimulating discussions, solicitations and contributions around the central theme of ALEAS, the development of adaptive learning systems in the field of Higher Education as a complementary tool for traditional courses, and promoting a community of practice in this field.
The volume collects 12 papers reporting reflections and quantitative studies covering mainly three topics: the assessment of the effects of anxiety or more generally of a different attitude in the study of Statistics, tools and methods for the assessment of training paths and technology-based learning experiences.

**Cristina Davino** is Associate Professor of Statistics (SECS-S/01) at the Department of Economics and Statistics, University of Naples Federico II. Her areas of expertise, in teaching and research, mainly concern multidimensional data analysis, statistical models and the construction of composite indicators applied to real contexts such as the assessment of quality of life and the analysis of learning processes.

**Rosa Fabbricatore** is a Psychologist and a PhD student in Social Sciences and Statistics at University of Naples Federico II. Her research interests focus on statistical methods applied to social sciences, educational and social health psychology. In particular, her main research topics relate to latent variable models such as structural equation models (SEM), Item Response Theory (IRT) models and Gaussian mixture models (GMM).

**Anna Parola** is a Psychologist and a Postdoc research fellow at the Department of Humanities, University of Naples Federico II. Her research areas include the difficulties of school-to-work transitions, career development, and career counseling. She is particularly interested in adolescents and youth career choices and the personal and environmental factors influencing learning processes and career paths.

ISBN: 978-88-6887-124-6
DOI: 10.6093/978-88-6887-124-6