



Adaptive Systems for foreign languages learning: towards a new tool in Canton Ticino (Switzerland)

Loredana Addimando¹, Elisa Rubegni², Francesca Mangili³, Alessandro Antonucci⁴, D'Alessio Germana⁵, Jan Hardie⁶

¹Department of Teaching and Learning, ²Department of Business Economics, Health and Social Care ; ³Department of Business Economics, Health and Social Care; ⁴Department of Business Economics, Health and Social Care; ⁵Department of Business Economics, Health and Social Care; ⁶ Department of Business Economics, Health and Social Care
University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

INTRODUCTION

Over the last decades the assessment of language knowledge has been done through the use of computers and, from '90s, we start to hear about Computer-Based Testing (CBT) and Computer Adaptive Testing (CAT). CBT and CAT have significantly impacted general assessment methods, especially in terms of large-scale group evaluation, because of their significant advantages over traditional paper-and-pencil tests. As mentioned by Chalhoub-Deville et al. (1999), CBT has remarkable potential benefits such as individual administration of tests, reduced pressure in terms of scheduling and supervising tests, more frequent and cheaper testing; greater standardization of measures, collection and storage of various types of information about test takers responses (e.g., response time, item review strategies, items omitted or not reached, etc.); possibility of special accommodations for test takers with disabilities, etc.

The language tests currently used by The University of Applied Sciences and Arts of Southern Switzerland (SUPSI) are classical CBT tests, thus making all these potential advantages possible. CAT would be a further development, aimed at improving the accuracy of the evaluation, as it is based on the delivery of questions based on the student's previous answers, and on a continuous tuning of the level of difficulty of the questions. As highlighted by Green (1983), CAT has many additional benefits such as improved test security, by providing large banks of questions, making it extremely difficult for test takers to learn the answers to all of the questions; rate of response can be used as an additional way of assessing the ability of a test taker; all test takers are challenged, and none are discouraged, as most questions are near their ability level, etc.

However, the successful implementation of CAT shouldn't been limited by a number of technical and practical aspects. As Brown (1994) expressed it, the acquisition of a new language is a fascinating, though colossal, enterprise, encompassing a wide range of variables that may stem from neurological to psychological, cognitive, and affective domains. The two main learning domains are defined, respectively, as "the mental side of human behavior", and "the emotional side of human behavior" (p. 135). Specifically, in affective domain, negative emotions and anxiety levels seem to be strongly implicated in a language test performance (Birjani & Alemi, 2015).

AIM & METHODOLOGY

The research project is aimed at developing the Adaptive Language Test (ALT): a tool for measuring foreign language competencies in academic contexts. In order to achieve this result a new "pilot test" was created and administered to a sample of students (n = 291). Resulting data were compared to the students' performance to a traditional test.

Along with the development of the adaptive test, authors were interested in understanding the whole experience of the student and, specifically, the respondents' perception facilitating or hindering the performance on the test. Among the psychological variables considered, anxiety and the perception of effectiveness/efficiency of the system were explored.

The aim of the present study is twofold:

- 1) To build a model able to exploring the link between performance and psychological aspects.
- 2) To verify the evaluated the effect of anxiety comparing the adaptive and the non-adaptive online test.

For this reason at the end of the test, each student was asked to fill out a short questionnaire about the experience they made. The measures included the constructs of anxiety, measured by the STAI (State - Trait Anxiety Inventory, Spielberg, 1989) and usability (with three sub-domains: efficacy, usability and involvement), measured by an *ad hoc* questionnaire.

In order to test the reliability and validity of the *ad hoc* tool, preliminary Explorative Factor Analysis (EFA) were conducted. Data cleaning procedures revealed a very low level of missing values (less than 1%) as well as the absence of multivariate outliers (Mahalanobis' distance was set to $p < .001$). Statistical indicators of robustness were: KMO and Bartlett test. The criterion to retain the most appropriate number of factors was Kaiser rule, eigenvalue > 1.00 and Cattell's scree test. The rotation was set to Oblimin and saturation values of each item were analysed. In this case, both KMO test (KMO=.751) and the Bartlett's test ($\chi^2=16.07$, $p > .001$) confirmed the opportunity to perform an EFA.

With regards to the Structural Equation Modeling (SEM), a set of both absolute and relative goodness of fit indexes were calculated. Specifically, we adopted χ^2 , Normed-Chi Square (NC), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual, Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) (see Hu & Bentler, 1999 for details). Model fit was considered robust if TLI was .95 (Cheung & Rensvold, 2002). Finally, convergent validity was tested administering the Bulatz and Oxford tests (two currently used commercial tests). These kind of procedures provide useful information about the potential generalizability of an adaptive test (Hair, Black, Babin, & Anderson 2010).

RESULTS

The correlations show associations in the context of interest (see table 1).

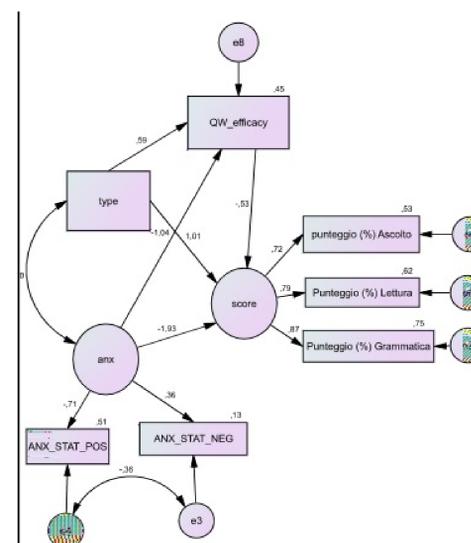
Table 1.

Correlations Matrix (in bold $p < .01$)

	1	2	3	4	5	6	7	8	9	10
QW_efficacy (1)	-									
QW_usability (2)	-.260	-								
QW_involvement (3)	.253	-.017	-							
ANX_STAT_POS (4)	.347	-.297	.068	-						
ANX_STAT_NEG (5)	-.131	.379	.102	-.434	-					
ANX_TRAIT_POS (6)	.160	-.077	.105	.536	-.220	-				
ANX_TRAIT_NEG (7)	-.045	.108	-.060	-.300	.330	-.552	-			
Listening (8)	.224	-.235	-.027	.368	-.213	.159	-.092	-		
Reading (9)	.223	-.188	-.079	.359	-.213	.167	-.093	.573	-	
Grammar (10)	.192	-.242	-.094	.383	-.237	.169	-.123	.597	.648	-

Analysis of the correlations among target variables is in line with the literature evidences and indicated that scores on Listening (8), Reading (9) and Grammar (10) are positively correlated with the positive state of anxiety (ANX_STAT_POS) and negatively correlated with negative state of anxiety (tab.1). The efficacy (measured by items such as "the questions was clear", "the time for answering was enough", etc.) positively correlated with listening ($r=.224$); reading ($r=.223$); and grammar ($r=.192$). On the contrary, the usability (measured by items such as "I've been able to explore the test online (i.e. Access the home screen, enter my data, go back, go forward, delete ..)", etc.) results negatively correlated with the target variables (respectively, listening, $r=-.235$; reading, $r=-.188$; and grammar, $r=-.242$).

In order to analyze the relations among the target variables, a Structural Equation Model (SEM) was tested. The measurement model are reported in figure 1. The SEM results suggest a good statistical significance ($\chi^2=16.98(8)$, $p > .75$) and a great practical significance (RMSEA = .49, TLI = .98, CFI = 0.99).



Results of the SEM were in line with the expected directions and magnitude of relationships among considered theoretical constructs. A more in-deep analysis shows that there is a direct negative effect of the anxiety on the performance (score). Furthermore, as an expected results, model shows a direct negative effect also between anxiety and the subjective perception of test efficacy. In other words, seems that anxiety is a predictor of the test efficacy. Anxious students evaluated the adaptive test efficacy less than students with low score of anxiety.

CONCLUSION

Given the growing interest in the CAT, the availability of an ALT should be of great value in monitoring students' performance in academic contexts. In addition, the opportunity to gather reliable information of anxiety statement among students contributed to the process of knowledge about the "the mental side of human behavior", and "the emotional side of human behavior" (*ibidem*, p.135). Finally, the questionnaire supported the notion of a direct influence of the anxiety on the students' performance more intensive with the adaptive system.

The explanation of this phenomenon could be due to greater interaction with the machine (more stimuli during the test, such as "informative popup" and so on), further investigations are actually in progress.

BIBLIOGRAPHY

- Birjandi, P. & Alemi, M. (2015). The Impact of Test Anxiety on Test Performance among Iranian EFL Learners. *BRAIN - Broad Research in Artificial Intelligence and Neuroscience*, 1(4), 43-58.
- Brown, H. D. (1994). *Teaching by Principles: an Interactive approach to Language Pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Chalhoub-Deville, M., Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Hair, J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper saddle River, New Jersey: Pearson Education International.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Spielberger, C. D. (1989). *State-Trait Anxiety Inventory: A comprehensive bibliography*. Palo Alto, CA: Consulting Psychologists Press