Papamitsiou, Z., Terzis, V. and Economides, A. A., (2014), Temporal learning analytics for Computer-based testing, 4th Conference in Learning Analytics and Knowledge (LAK14), Indianapolis, USA

Temporal Learning Analytics for Computer Based Testing

Zacharoula K. Papamitsiou
Interdepartmental Progr. of Postgraduate Studies in Information Systems, University of Macedonia
156 Egnatia Avenue, Thessaloniki, 54636, Greece
+30 2310 891-768
papamits@uom.edu.gr

VasileiosTerzis
Interdepartmental Progr. of Postgraduate Studies in Information Systems, University of Macedonia
156 Egnatia Avenue, Thessaloniki, 54636, Greece
+30 2310 891-768
bterzis@otenet.gr

Anastasios A. Economides
Interdepartmental Progr. of Postgraduate Studies in Information Systems, University of Macedonia
156 Egnatia Avenue, Thessaloniki, 54636, Greece
+30 2310 891-768
economid@uom.gr

ABSTRACT
Predicting student's performance is a challenging, yet complicated task for institutions, instructors and learners. Accurate predictions of performance could lead to improved learning outcomes and increased goal achievement. In this paper we explore the predictive capabilities of student's time-spent on answering (in-)correctly each question of a multiple-choice assessment quiz, along with student's final quiz-score, in the context of computer-based testing. We also explore the correlation between the time-spent factor (as defined here) and goal-expectancy. We present a case study and investigate the value of using this parameter as a learning analytics factor for improving prediction of performance during computer-based testing. Our initial results are encouraging and indicate that the temporal dimension of learning analytics should be further explored.

Categories and Subject Descriptors
D.4.8[**Performance**]: Modelling and prediction; K.3.1 [**Computer Uses in Education**]: Predictive applications of data

General Terms
Performance, Experimentation, Measurement, Prediction, Computer based assessment, Educational Data Mining.

Keywords
Computer-based testing, temporal learning analytics, goal-expectancy, prediction of performance

1. INTRODUCTION
Learning Analytics (LA) is an ecosystem of methods and techniques (in general procedures) that successively gather, process, report and act on machine-readable data on an ongoing basis in order to reflect on learning processes [1]. Like any other context-aware system, the LA procedures monitor, track and record data related to the context, interpret and map the real current state of these data, organise these data (e.g., filter, classify, prioritise), use these data (e.g., decide adaptations, recommend, provide feedback, guide the learner) and predict the future state of these data [2]. The target is to inform and empower learners, instructors and organization about performance and goal achievement, and facilitate decision making accordingly.

Prediction of performance is a challenging task for institutions, instructors and learners. Accurate and up-to-date predictions could have significant effect on strategic adjustments that could lead to improved learning outcomes and increased goal achievement. A number of case studies explore, identify and evaluate factors as indicators of performance that are more significant for prediction purposes. These factors include demographic characteristics, grades (either on course assignments or final exams scores), students' portfolios, students' participation, enrollment and engagement in activity and students' mood and affective states [3, 4, 5, 6, 7, 8, 9].

Research shows a significant positive relationship between participation and grades [10]. Performance is also related to the type, content and nature of both formative and summative feedback students receive during formative assessment [7, 11].

Another possible factor for predicting performance, yet not extensively explored in literature, is the temporal dimension of students' engagement in activity. Xiong et al. [12] applied machine learning and statistical analysis to examine the effect of students' response time on the prediction of their performance and how response time could be used.

In a former study [13] the authors examined the results of students' time-spent studying on their academic performance and evaluated the interaction of motivation with study time. The results were encouraging and shown a relationship between motivation and study time. Additionally, Beal and Cohen [14] experimented with modeling "the amount of time students are willing to spend on problems and how it changes over the course of a session" (pg. 67). In their work, the authors modeled unobservable factors such as engagement, using hidden variable models.

Furthermore, in the context of Computer Adaptive Testing (CAT), the orientation information includes time among others [15]. In this case, the goal is the adaptation engine to apply the adaptation with respect to the orientation information, in order to support the examinee. The author gives examples of how and what the engine should adapt regarding orientation to time.

LAK '14, March 24 - 28 2014, Indianapolis, IN, USA

http://dx.doi.org/10.1145/2567574.2567609

Within this paper we explore the predictive capabilities of student's time-spent on answering (in-)correctly each question of a multiple choice assessment quiz, along with student's final quiz-score. Our goal is to investigate whether time-spent on correct answers and time-spent on wrong answers could be formalized as a predictive model that explains the actual performance during computer-based testing. We define this procedure as "temporal learning analytics". Furthermore, we discovered a relationship between temporal learning analytics and student's goal expectancy. In our work, we present the results from a case study with 96 participants, as well as future research dimensions.

The rest of this paper is organized as follows: in section 2, we present our experiment methodology. In section 3, we analyze the results from the case study. In section 4, we discuss about our findings, share our initial conclusions and describe our future work plans.

2. METHODOLOGY

2.1 Experiment Process and Data Collection

For our case study we used a simplified version of the LAERS assessment environment [1]. We developed a new module that consists of two components:

a) a computer-based testing unit in multiple choice quiz format, and

b) a tracker that logs students' activity data.

The testing unit displays the multiple choice quiz with a predefined (by the instructor), stable (for all examinees) number of questions and of fixed duration. Each question is displayed separately and one-by-one. During the quiz, the student can skip a question (either because he/she is not sure about the answer, or because he/she thinks it is difficult), and answer it later. The list of skipped questions is displayed alongside the quiz, within the same window. The student temporarily saves his/her answers on the quiz questions, before finalizing his/her decision. The student can also change his/her initial choice, and save a new answer. He/she submits the quiz answers only once, whenever estimates that he/she is ready to do so, within the duration of the exam. In case a student chooses not to submit an answer to a question, he/she receives zero points for this question. Students' submitted answers are stored on a database.

The second component records students' activity data during testing. In specific, the tracker logs for each student the following parameters:

- **quest_id**: the question the student works on,
- **ans_id**: the answer the student submits on a question,
- **rw**: the correctness of the submitted answer (right or wrong),
- **count_view**: how many times the student views each question,
- **count_changes**: how many times the student changes the answer he/she submits for each question,
- **idle_t_view**: the time the student spends on viewing each question (not saving an answer),
- **t_ans**: the time the student spends on answering each question (saves an answer).

Figure 1 illustrates the student's view of the environment during testing.

Figure 1. The LAERS environment during testing

The system also calculates a) the Final Score (that is the Actual Performance - **AP**) for each student, and b) average time values for each question and each student. Students' data and quiz's data are stored on a database, but tracked activity data are stored on a log file (results.csv).

We also embedded into the system a pre-test questionnaire in order to measure each student's goal expectancy. Data from the questionnaire were logged on a separated file (pretest.csv).

Data were collected from a total of 96 participant students of a European High School, aged 16 years old. 9 groups of 10 to 12 students attended the midterm exams, for 30 minutes each group, from 2nd to 7th of October 2013. The 12 multiple choice questions of the test were related to the basic concepts of Computer Science. The final log file (results.csv) contained 4133 rows of raw data.

2.2 Research model and Hypotheses

Based on previous studies, this paper goes a step further by introducing time and students' perceptions as significant predictors of actual performance.

Total_time_answer_correct (TTAC):

We define TTAC as the total time that a student spends on viewing the questions and submitting the correct answers. We believe that a student who spends more time for choosing the correct answers is more likely to have better performance. In specific, a student that answers correctly many questions will aggregate more and more time to TTAC. Therefore, we hypothesized:

H1: *TTAC will have a positive effect on Actual Performance*

Total_time_answer_wrong (TTAW):

We define TTAW as the total time that a student spends on viewing the questions and submitting the wrong answers. We believe that a student who spends more time for choosing the wrong answers is more likely to have lower performance. In specific, a student that answers incorrectly many questions will aggregate more and more time to TTAW.

Thus, we hypothesized:

H2: *TTAW will have a negative effect on Actual Performance*

The total time is the aggregation of the above two time units.

Goal Expectancy (GE)

A variable which measures self-confidence and goal orientation regarding the use of a Computer Based Assessment is Goal Expectancy (**GE**), which was proposed in Computer Based Assessment Acceptance Model (CBAAM) [16]. GE has two dimensions. The first dimension is student's preparation to take the CBA. GE actually measures if a learner is fulfilled with his/her preparation. The second dimension includes the desirable level of success for each student. The students, before taking the CBA, set a goal regarding a percentage of correct answers that provides them a satisfying performance. We believe that GE is highly positively correlated with the TTAC. The reason is that a well prepared student will answer more questions correctly; therefore the time that he/she will spend for the correct answers will be higher than the spending time of a poorly prepared student. On the contrary, a well prepared student will have fewer wrong answers than a poorly prepared student. Consequently, a well prepared student will spend less time on questions that finally will answer wrong. Therefore, GE is expected to have a negative impact on TTAW.

Thus, we hypothesized:

H3: *GE will have a positive effect on Total Time to answer correct*.

H4: *GE will have a negative effect on Total Time to answer wrong*.

To summarize, this paper develops and explores a causal model to determine student's Actual Performance (Figure 2).

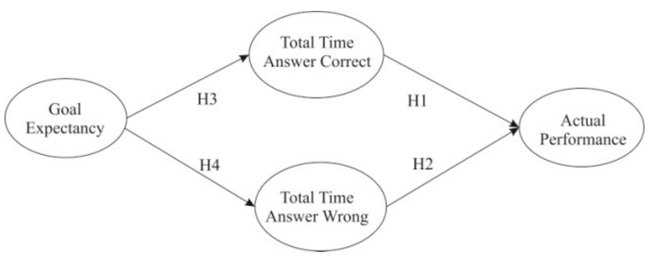

Figure 2. Research Model

2.3 Measures

In order to examine the constructs of the model, we used data collected with modified LAERS [1], and we also adapted three items from the questionnaire to measure Goal Expectancy [16]. In specific, these three items are:

- GE1: Courses' preparation was sufficient for the CBA
- GE2: My personal preparation for the CBA
- GE3: My performance expectations for the CBA

We used the technique of partial least-squares (PLS) analysis to evaluate the measurement and the structural model. Previous studies supported that PLS is a powerful tool to develop and test theories in early stages, and to predict with small samples [17, 18, 19].

PLS follows two guidelines regarding the sample size. The first is that the sample has to be 10 times larger than the number of items for the most complex construct. The second is that the sample has to be 10 times the largest number of independent variables impacting a dependent variable [19]. The most complex variable of the proposed model is GE with three items. Therefore, the sample of 96 participants surpassed the recommended value of 30.

Reliability and validity of the measurement model are proved by measuring the internal consistency, convergent validity and discriminant validity [20]. In our proposed model the measurement model analysis is necessary mainly for GE which is the only latent variable. More specifically, a value higher than 0.7 is acceptable regarding the items' factor loadings on the corresponded constructs. In order to analyze discriminant validity, we have also to examine AVE (Average Variance Extracted). AVE should be higher than 0.5 and the AVE's squared root of each variable should be larger than any correlation with every other construct [18, 19, 20]. Finally, Composite reliability and Cronbach alpha should be also examined. Composite reliability and Cronbach alpha are considered acceptable when they scored over 0.7 [21, 22].

The structural model and hypotheses are examined mainly by two criteria:

(1) by evaluating the variance measured for (R^2) by the antecedent constructs. Previous studies suggested 0.2, 0.13 and 0.26 as small, medium and large variance respectively [23];

(2) the significance of the path coefficients and total effects by using bootstrapping procedure and calculating the t-values.

In order to examine the measurement and the structural model we used SmartPLS 2.0 [24].

3. Results

Table 1 confirms the adequate values for the measurement model.

Table 1. Results for the Measurement Model

Construct Items	Factor Loading (>0.7)[a]	Cronb. a (>0.7) [a]	C.R. (>0.7) [a]	AVE (>0.5) [a]
GE		0.74	0.84	0.65
GE1	0.78			
GE2	0.77			
GE3	0.85			
TTTAC	1.00	1.00	1.00	1.00
TTAW	1.00	1.00	1.00	1.00
AP	1.00	1.00	1.00	1.00

[a] Indicates an acceptable level of reliability and validity

In addition, Table 2 presents the correlation matrix. The diagonal elements are the square root of the average variance extracted (AVE) of a construct. Discriminant validity is confirmed since the diagonal elements are higher than any correlation with another variable.

Table 2. Discriminant validity for the measurement model

Construct	GE	TTAC	TTAW	AP
GE	**0.8**			
TTAC	0.35	**1**		
TTAW	0.34	-0.47	**1**	
AP	0.37	0.74	0.58	**1**

A bootstrap procedure with 1000 resamples was used to test the statistical significance of the path coefficients in the model. The results for the hypotheses are summarized in Table 3 and illustrated in Figure 3. TTAC and TTAW have significant direct positive and negative effect on AP respectively. Moreover GE is a determinant of TTAC and TTAW as well. Thus all the hypotheses were confirmed.

Table 3. Hypothesis testing results

Hypothesis	Path	Path coeff.	t value	Results
H1	TTAC -> AP	0.60*	8.6	support
H2	TTAW ->AP	-0.29*	3.3	support
H3	GE ->TTAC	0.35*	4.5	support
H4	GE ->TTAW	-0.23*	2.83	support

*p<0.01

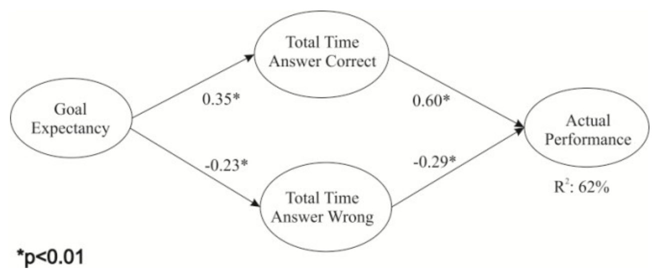

Figure 3. Path coefficients of the research model

Additional to the direct effects, the structural model includes also indirect effects (Table 4).

Table 4. R^2 and Direct, Indirect and Total effects

Dependent Variable	R^2	Independent Variables	Direct effect	Indirect effect	Total effect
AP	0.62	TTAC	0.60	0.00	0.60*
		TTAW	-0.29	0.00	0.29*
		GE	0.00	0.28	0.28*

Moreover, in the PLS analysis the R^2 values are used as a goodness-of-fit measure [25]. The model explains almost the 62% of the variance in AP. The total effects of TTAC (0.60), of TTAW (-0.29) and GE (0.28) are strong.

4. DISCUSSION, CONCLUSIONS AND FUTURE WORK

The principal idea of this work is to introduce the temporal data regarding the development of more personalized and fully automated systems. In this paper we explored the predictive capabilities of temporal learning analytics regarding Actual Performance a student achieves during computer based testing. The temporal interpretation of students' performance in activity, attempted here, could be used for predicting their progress. In other words, interpreting students' participation and engagement in terms of "time-spent" could lead to a complementary dimension of a more concise predictive model.

For the needs of our experiment we configured the LAERS system and implemented additional components. 96 students from Secondary Education participated in our case study, in order to estimate the validity of our hypotheses. Next, we used the PLS technique to evaluate the measurement and the structural model.

Our initial results indicate three main contributions: a) a direct positive effect of total time to answer correct on AP, b) a direct negative effect of total time to answer wrong on AP, and c) an indirect effect of GE on AP.

In specific, our preliminary results highlight a detected trend that temporal learning analytics have a statistically significant capability on predicting actual performance (almost 62% of the variance). In particular, TTAC has a direct positive effect (0.60) on AP. This means that if a learner spends more time to answer correctly, it is more likely to score higher. Moreover, TTAW has a direct negative effect (-0.29) on AP. This means that if a learner spends more time to answer incorrectly, it is more likely to score lower.

In case this finding is verified by further experiments, it could indicate that temporal learning analytics could replace traditional grading in assessment activities where grades act as effort overload. In these activities, an "hourglass" visualization (indicating real-time TTAC/TTAW) could be embedded into the CBA system to inform learners (and instructor) about their progress and performance.

Furthermore, the third finding has two additional inherent dimensions: a) there is a positive effect of GE on TTAC, and b) there is a negative effect of GE on TTAW. Our model demonstrates that GE is a direct strong determinant of the temporal variables and concurrently an indirect strong determinant of AP (Table 4).

Although someone would expect such results, there are no other research studies that provide evidence. Our findings indicate that the temporal dimension of student's participation in assessment activities could be acknowledged as a predictive factor of performance.

However, since this research is one of the first that introduce temporal dimension to predict performance, further research should be applied to verify our results.

As stated on section 2.2, Goal Expectancy is a dimension of student's perceptions, and in the context of this case study, GE is about the student's perception of preparation. It would be very interesting to investigate the correlation between student's perception of performance and his/her actual performance. It would be important to detect the temporal factors that determine student's perception of performance. Further, these factors could be measured and embodied in the structural model as additional parameters.

In addition, students' self-confidence and certainty during testing constitute two parameters that could also be explored through their temporal instantiation. It would be interesting to examine the relationships between these behaviors and the total idle time a student spent on a question, or how many times he/she skipped a question.

Further, the proposed ideas should be examined and combined with other features such as learner's characteristics (e.g. gender, age, personality traits) and different infrastructures (e.g. universities) [2, 26, 27].

To conclude, we believe that this approach could be exploited for the construction of more sophisticated CBA systems that record the temporal behavior of their users, analyze this information and adapt their functionalities accordingly, in order to improve the users' performance.

5. REFERENCES

[1] Papamitsiou, Z. & Economides, A. A. 2013. Towards the alignment of computer-based assessment outcome with learning goals: the LAERS architecture, *In Proc. of the IEEE Conference on e-Learning, e-Management and e-Services (IC3e 2013)*, Malaysia

[2] Economides, A. A. 2009. Adaptive context-aware pervasive and ubiquitous learning. *International Journal of Technology Enhanced Learning*, 1(3), pp. 169-192.

[3] Abdous, M., He, W., & Yen, C.-J. 2012. Using data mining for predicting relationships between online question theme and final grade. *Educational Technology & Society*, 15 (3), pp. 77–88.

[4] Fancsali, S. E. 2011. Variable construction for predictive and causal modeling of online education data. *In Proc. of the 1st International Conference on Learning Analytics and Knowledge (LAK '11)*. ACM, New York, USA, pp. 54-63.

[5] Huang, C-J., Chu, S-S., & Guan, C-T. 2007. Implementation and performance evaluation of parameter improvement mechanisms for intelligent e-learning systems. *Comput. Educ*. 49(3), pp. 597-614

[6] Pardos, Z. A., Baker, R. S. J. D.,San Pedro,M. O. C. Z.,Gowda,S. M. & Gowda, S. M. 2013. Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes. *In Proc. of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, ACM, New York, USA, pp. 117-124.

[7] Tanes, Z., Arnold, K.E., King, A. S. & Remnet, M.A. 2011. Using Signals for appropriate feedback: Perceptions and practices. *Comput. Educ*. 57(4), pp. 2414-2422.

[8] Wolff, A., Zdrahal, Z., Nikolov, A. & Pantucek, M. 2013. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *In Proc. of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, ACM, New York, USA, pp. 145-149.

[9] Moridis, C. N. & Economides, A. A. 2009. Prediction of student's mood during an online test using formula-based and neural network-based method. *Comput. Educ*, 53(3), pp. 644-652.

[10] Macfadyen, L. P, & Dawson, S. 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Comput. Educ*. 54, 2, pp. 588-599.

[11] Wilson, K., Boyd, C., Chen, L. & Jamal, S. 2011. Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Comput. Educ*. 57(2), pp. 1493-1500.

[12] Xiong, X.,Pardos, Z. & Heffernan, N. 2011. An analysis of response time data for improving student performance prediction, *In proc. KDD 2011 Workshop: Knowledge Discovery in Educational Data*, Held as part of 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining in San Diego

[13] Nonis, S. A., & Hudson, G. I. 2006. Academic Performance of college students: influence of time spent studying and working, *Journal of Education for Business*, 81(3).

[14] Beal, C.R. & Cohen, P. R. 2008. Temporal data mining for educational applications. In *Proc. of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence* (PRICAI '08), Springer-Verlag, Berlin, Heidelberg, pp. 66-77.

[15] Economides, A. A. 2005. Adaptive orientation methods in computer adaptive testing. In *Proc. E-Learn 2005 World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 1290-1295.

[16] Terzis, V., & Economides, A. A. 2011. The acceptance and use of computer based assessment. *Comput. Educ.,* 56(4), pp. 1032–1044.

[17] Falk, R. F., & Miller, N. B. 1992. *A Primer for Soft Modeling*. Akron, OH: University of Akron Press.

[18] Fornell C., & Bookstein F.L. 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19, pp. 440–452.

[19] Chin, W. W. 1998. *The partial least squares approach to structural equation Modeling*, in Marcoulides G. A. (eds.) Modern Business research Methods (pp. 295 -336). Mahwah, NJ: Lawrence Erlbaum Associates.

[20] Barclay, D., Higgins, C., & Thompson, R. 1995. The Partial Least Squares approach to causal modelling: Personal computer adoption and use as an illustration. *Technology Studies*, 2(1), pp. 285-309.

[21] Agarwal, R. & Karahanna, E. 2000. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 24, pp. 665-694.

[22] Compeau, D., Higgins, C.A., & Huff, S. 1999. Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 23, pp. 145–158.

[23] Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: 2nd ed., Erlbaum.

[24] Ringle, C. M., Wende, S., & Will, A. 2005. SmartPLS 2.0 (beta). University of Hamburg, Germany, http://www.smartpls.de

[25] Hulland, J. 1999. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal*, 20(2), pp. 195–204.

[26] Terzis, V., & Economides, A. A. 2011. Computer based assessment: Gender differences in perceptions and acceptance. *Computers in Human Behavior*, 27(6), pp. 2108-2122.

[27] Terzis, V., Moridis, C. N., & Economides, A. A. 2012. How student's personality traits affect Computer Based Assessment Acceptance: Integrating BFI with CBAAM. *Computers in Human Behavior*, 28(5), pp. 1985-1996.