STUDENTS' PERCEPTION OF PERFORMANCE VS. ACTUAL PERFORMANCE DURING COMPUTER BASED TESTING: A TEMPORAL APPROACH

Zacharoula Papamitsiou, Anastasios A. Economides

Interdepartmental Progr. of Postgraduate Studies in Information Systems, University of Macedonia (GREECE)

Abstract

In this paper we present a case study of tracking a) students' perceptions of performance and goal expectancy before taking a computer-based test, b) their perception of performance after taking the test, c) their actual performance as it is calculated by the testing environment itself and d) their time-spent behavior during test. Our goal is to explore whether students' time-spent behavior during computer-based testing – expressing "(un-)certainty" - can reveal any differences between what they believe they know, and what they actually know. Furthermore, we investigate the correlation between students' goal-expectancy and their "(un-)certainty". We conducted a case study with a simplified version of the LAERS assessment environment. We used statistical methods and Structural Equation Modeling (SEM) for the construction of a predictive model which explains the results, based on students' time-spent on answering each question of a multiple choice quiz. Initial results indicate that a) students' perceptions of performance and their actual performance significantly differ, both pre and post test, b) students' temporal behavior can explain satisfactorily what they actually know, and c) goal-expectancy has an indirect effect on students' "(un-)certainty".

Keywords: computer-based assessment, computer-based testing, goal expectancy, LAERS, learning analytics, prediction of performance, temporal behavior, (un-)certainty.

1 INTRODUCTION

Assessment and performance are bidirectionally interrelated. Processing large amounts of gathered educational data have potential to clearly determine and evaluate what students already know and set the boundary between that and what they need to learn. In [1], the authors set under discussion the issue of considering and modeling students' perceptions when dealing with performance. Perception of performance refers to what students believe they know. Keren [2] defines the degree to which a person's perception of performance corresponds to his/her actual performance as "calibration". Students' beliefs of what they have learned as well as their goal-setting are important because they reflect on students' effort, (self-)awareness and achievement-related behaviours. Most studies correlate students' perceptions of performance with self-confidence, task-difficulty and motivation [3], [4], [5], [6], [7], [8].

In literature, there have been many case studies that explore, identify and evaluate factors as indicators of performance for prediction purposes. Among these, demographic characteristics, grades (either on course assignments or final exams scores), students' portfolios, multimodal skills, students' participation, enrollment and engagement in activity and students' mood and affective states are acknowledged as the most significant ones [9], [10], [11], [12], [13], [14].

Beyond these factors, the temporal dimension of students' engagement in activity is also under investigation about its predictive capabilities. Researchers examine both the effect of students' response time [15], the students' time-spent studying regarding the interaction of motivation with study time [16] and "the amount of time students are willing to spend" on problem-solving [17, pg. 67]. In [1] the authors investigated whether time-spent on answering (in-)correctly could be formalized as a predictive model ("temporal learning analytics") that explains the actual performance during computer-based testing.

In this paper we present a case study of tracking a) students' perceptions of performance and goal expectancy before taking a computer-based test, b) their perception of performance after taking the test, c) their actual performance as it is calculated by the testing environment itself and d) their time-spent behavior during test. Our goal is to explore whether students' time-spent behavior during computer-based testing – expressing "(un-)certainty" – can reveal any differences between what they

ISBN: 978-84-616-8412-0

believe they know, and what they actually know. Furthermore, we investigate the correlation between students' goal-expectancy and their "(un-)certainty". We extend the previously defined "temporal learning analytics" procedure [1] with the "(un-)certainty" parameter. We present the results from a case study conducted with Secondary Education participants, discuss our findings and propose future research directions.

The rest of this paper is organized as follows: in section 2, we present our experiment methodology, the data collection procedure and the research hypotheses. In section 3, we analyze the results and propose our extension of the predictive model (suggested in [1]). Finally, in section 4, we discuss about our findings, share our conclusions and highlight our plans on future work.

2 METHODOLOGY

2.1 The testing environment

For our case study we used a simplified version of the LAERS assessment environment [18]. We developed a new module that implements a testing mechanism. In specific, this module consists of two components:

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

b. a tracker that logs students' activity data.

For the purpose of our study, we wanted to measure the students' time-spent behavior on each question. We wanted to allow the students to freely arrange the order of answering the questions. For that reason, we chose to deliver each question separately one-by-one and in random order.

The testing unit displays the multiple choice quiz with a predefined (by the instructor), stable (for all examinee) number of questions and is of fixed duration. The student can temporarily save his/her answers on the quiz questions, before finalizing his/her decision. He/she submits the quiz answers only once, whenever estimates that he/she is ready to do so, within the duration of the exam. Students' submitted answers are stored on a database.

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 can also change his/her initial choice, and save a new answer. In case a student chooses not to submit an answer to a question, he/she receives zero points for this question.

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

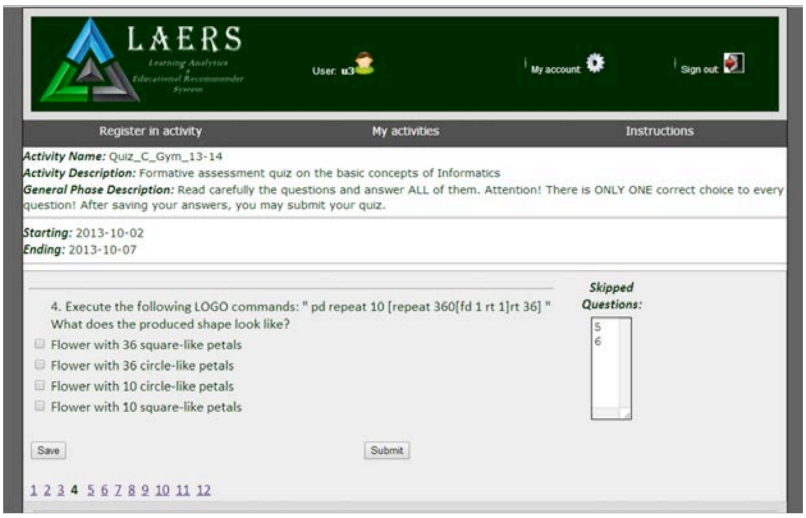

Fig. 1. The LAERS environment during testing

The system also computes the final score for each student (that is the Actual Performance - **AP**). This score is calculated only once the student decides to submit the quiz, and its value is stored on the database.

The second component – the tracker – records students' activity data during testing on a log file. In particular, the following variables are tracked (Table 1):

Table 1. Tracked variables and their description

Variable	Description
quest_id	the question the student works on
ans_id	the answer the student saves (or submits) on a question (every time he/she saves an answer)
rw	the correctness of the saved (or 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 saves for each question
idle_t_view	the time the student spends on viewing each question
t_ans	the time the student spends on answering each question

The system also calculates average time values for each question and each student, as well as average time a student changes his/her answers.

2.2 Data Collection

For the purpose of our case study, we also embedded into the system two questionnaires: a pre-test and a post-test questionnaire, in order to measure each student's goal expectancy (pre-test) and record perception of performance (pre-test and post-test). Data from the questionnaires were logged on two separated files (pretest.csv and posttest.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 Informatics. All questions used in the current case study correspond to the lower three levels of the cognitive domain of Bloom's taxonomy (Remembering, Understanding and Applying) [19]. The final log file (results.csv) contained 4133 rows of raw data.

2.3 Research Hypotheses

2.3.1 Perception of performance

In general, perception of performance is a subjective estimation of what a student believe he/she has learned. This perception can be separated into two phases: one before taking an exam, and one after taking the exam. The discrimination between these two and what the student actually knows and how this perception changes during an exam is the question in this study.

Usually, when students are preparing themselves for an exam, they study the suggested educational material and assess what they have learned and how much satisfied they are from their preparation. Given a specific study material, they determine the desired level of achievement according to their effort on understanding and knowledge acquisition. However, parameters such as their psychological and emotional state, for example, may garble their sense of objectivity about what they have really learned. Thus, we hypothesized that:

RH1*: Pre-test students' perception of performance is significantly different from their Actual Performance*

Furthermore, after taking the exam, students' perception of performance is about what they believe they have achieved. In many cases it represents their satisfaction regarding their performance, their feeling of (in-)appropriate preparation, their perception of difficulty of the quiz and more. However, possible misconceptions, time-pressure, stress, suitability of the test or other factors, may also result to subjective judgement of their performance. Thus, we hypothesized that:

RH2*: Post-test students' perception of performance is significantly different from their Actual Performance*

2.3.2 (Un-)certainty and Goal Expectancy

Certainty is defined as “a subjective sense of conviction or validity about one’s attitude or opinion” [20, p. 215]. It is used to describe a person‘s strength of belief about the accuracy or quality of a mental representation, prediction, judgment, or choice [21], [22]. Thus, certainty can be described on a continuum ranging from total confidence to complete doubt. Lack of certainty (i.e., doubt) is often conceptualized as an inhibitor to the use of the construct one is uncertain about [23], [24].

In a sense, there are two opinions describing a student regarding his/her certainty: a) the confident – that is a student who is convinced about the correctness of the answer (“I’m sure I know the answer!”) and b) the skeptical – that is a student who wants to be absolutely sure (overcome his/her doubts) about the answer he/she submits (“I don’t want to submit a wrong answer by superficiality”).

Confidence and skepticism both reflect on student’s self-efficacy and goal expectancy. For example, a student who is well prepared and aims to achieve a high score, could either be confident or cautious. However, a wrongly estimated perception of self-performance could lead an over-confident student to misconceptions and thoughtlessness. On the other hand, a diligent student usually is self-aware and balances his/her goals to his/her abilities. Consistent to this assumption are the findings from several studies in literature [5].

We define here as “(un-)certainty” a latent variable that consists of two sub-parameters during using a Computer Based Assessment (CBA): a) the total times the student views a question and b) the total time he/she spends on the question remaining idle (does not save or submit an answer). This means that the more certain the student *wants to be* before answering a question, the more the idle time he/she spends on that question. Furthermore, the more certain the student *wants to be*, the more the times he/she re-views the questions. Thus, “(un-)certainty” is a measure of cautiousness during the assessment.

In particular, we believe that (un-)certainty will have a positive impact on Actual Performance. That is because a cautious student is more likely to score higher than a hasty student. Thus, we hypothesized:

***H1**: (un-)certainty will have a direct positive effect on Actual Performance*

Furthermore, a variable which measures self-confidence and goal orientation regarding the use of a CBA is Goal Expectancy (**GE**), which was proposed in Computer Based Assessment Acceptance Model (CBAAM) [25]. GE actually measures if a learner is fulfilled with his/her preparation. The students, before taking the CBA, set a goal regarding a percentage of correct answers that provides them a satisfying performance. In other words, they estimate their self-confidence regarding their study and the assessment.

In a previous study [1], the authors found that GE has a direct positive effect on Total Time to Answer Correct (TTAC) and a direct negative effect on Total Time to Answer Wrong (TTAW). Since GE measures student’s self-confidence before taking a CBA and (un-)certainty measures his/her cautiousness during assessment, we believe that GE will have an indirect effect on (un-)certainty. In current study we believe that both TTAC and TTAW are highly positively correlated with (un-)certainty. The reason is that the time a cautious student will spend for the answers (regardless of whether they are correct or wrong), and the times he/she re-views the questions will increase his/her certainty.

Thus, we hypothesized the following:

***H2:** TTAC will have a positive effect on (un-)certainty*

***H3:** TTAW will have a positive effect on (un-)certainty*

This paper suggests a causal model that explores the (un-)certainty parameter as a determinant of student’s Actual Performance, as shown in Fig. 2.

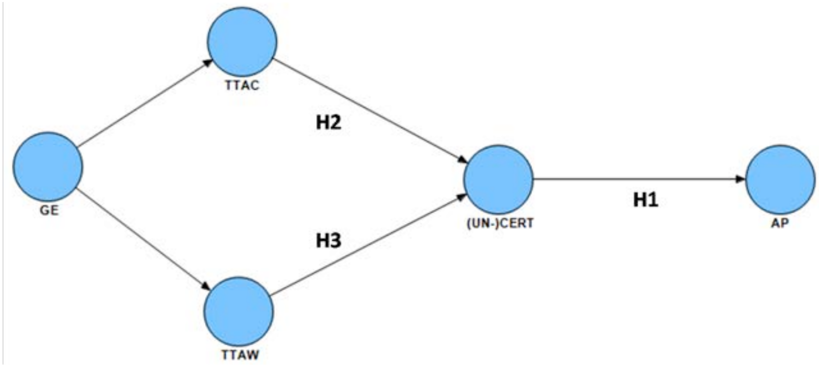

Fig. 2. Research Model

2.4 Measures

In order to explore the differences between perception of performance and actual performance we used one item from the pre-test questionnaire and one item from the post-test questionnaire.

The item representing perceived performance pre-test is:

- PPPre: What score do you believe you will get?

The item representing perceived performance post-test is:

- PPPost: What score do you believe you got?

We used Paired Samples t-test to examine if there is statistical difference between perceptions of performance and actual performance. Paired samples t-tests were conducted to compare perceived performance pre-test and actual performance, perceived performance post-test and actual performance and perceived performance pre-test and post-test.That is because: a) all variables are measured exactly the same way in the same unit at the same time and we wanted to examine if the participants maintain the same perceptions before and after the quiz, b) measurements were taken from the same participants before and after some manipulation (the quiz), and each value in every one of these samples has a natural partner in the other samples. The criteria for the statistical difference are the t-value and p-value.

Furthrmore, 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 [26], [27]. 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 most complex variable of the proposed model is GE with three items. Therefore, the sample of 96 participants surpassed the recommended value of 30. The second is that the sample has to be 10 times the largest number of independent variables impacting a dependent variable [27].

Reliability and validity of the measurement model are proved by measuring the internal consistency, convergent validity and discriminant validity [28], [29]. In our proposed model the measurement model analysis is necessary for (un-)certainty which is a 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 [26], [27], [28]. Finally, Composite reliability and Cronbach alpha should be also examined. Composite reliability and Cronbach alpha are considered acceptable when they scored over 0.7 [30], [31].

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 [32];

(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 use SmartPLS 2.0 [33].

3 RESULTS

3.1 Perceptions of performance vs. Actual Performance

Tables 2 (a) and 2(b) display the results from the Paired Samples t-tests. There was a significant difference in the scores for PPPre (M=16.91, SD=3.05) and AP (M=14.56, SD=4.73). There is strong evidence (t=5.53, p = 0.001) that students' perceptions of performance pre test are higher than their actual performance. In this data set, the divergence, on average, is approximately 2 points.

Furthermore, there was a significant difference in scores for PPPost (M=17.03, SD=2.95) and AP (M=14.56, SD=4.73). There is strong evidence (t=6.56, p = 0.001) that students' perceptions of performance post test are higher than their actual performance, even compared to their perceptions pre test. In this data set, the divergence, on average, is approximately 2,5 points.

However, paired samples t test failed to reveal a statistically reliable difference between the mean number of scores for PPPre (M=16.91, SD=3.05) and PPPost (M=17.03, SD=2.95). Since t=0.69, p=0.49, we can support that students' perceptions of performance pre test and post test remain approximately the same.

Table 2(a). Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PPPre	16,91	96	3,051	,311
	AP	14,56	96	4,726	,482
Pair 2	PPPost	17,03	96	2,945	,301
	AP	14,56	96	4,726	,482
Pair 3	PPPre	16,91	96	3,051	,311
	PPPost	17,03	96	2,945	,301

Table 3(b). Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	PPPre - AP	2,352	4,171	,426	1,507	3,197	5,526	95	,000
Pair 2	PPPost - AP	2,472	3,694	,377	1,723	3,220	6,556	95	,000
Pair 3	PPPre - PPPost	,120	1,707	,174	-,226	,466	,687	95	,493

3.2 The (un-)certainty effect

Table 3 confirms the adequate values (Factor Loadings, Cronbach alpha, Composite reliability and Average Variance Extracted) for the measurement model.

Table 3. Results for the Measurement Model

Construct Items	Factor Loading (>0.7)[a]	Cronbach Alpha (>0.7) [a]	Composite reliability (>0.7) [a]	AVE (>0.5) [a]
(un-)certainty		0.84	0.93	0.86
Total_Idle_time	0.92			
Total_ans_check_views	0.93			
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 4 presents the correlation matrix. The diagonal elements (Table 4) are the square root of the average variance extracted (AVE) of a construct. According to the Fornell-Larcker criterion [34], the AVE of each latent construct should be higher than the construct's highest squared correlation with any other latent construct. Discriminant validity is established when an indicator's loading on a construct is higher than all of its cross loadings with other constructs. Consequently, discriminant validity is confirmed since the diagonal elements are higher than any correlation with another variable.

Table 4. Discriminant validity for the measurement model

Construct	(un-)certainty	TTAC	TTAW	AP
(un-)certainty	**0.92**			
TTAC	0.54	**1**		
TTAW	0.10	-0.47	**1**	
AP	0.41	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 5 and illustrated in Fig. 3. TTAC and TTAW have significant direct positive effect on (un-)certainty respectively. Moreover (un-)certainty is a determinant of AP as well. Thus all the hypotheses were confirmed.

Table 5. Hypothesis testing results

Hypothesis	Path	Path coeff.	t value	Results
H1	(un-)certainty -> AP	0.41*	5.8	support
H2	TTAC ->(un-)certainty	0.68*	8.4	support
H3	TTAW ->(un-)certainty	0.31*	3.01	support

*p<0.01

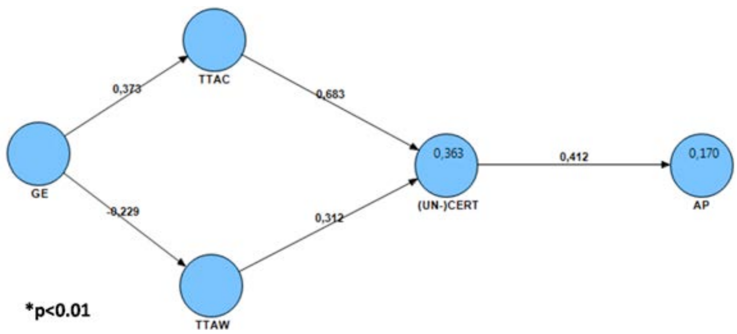

Fig. 3. Path coefficients of the research model

Additional to the direct effects, the structural model includes also indirect effects (Table 6). Specifically, the total effects of GE through TTAC and TTAW indicates that GE is also a determinant of (un-)certainty.

Table 6. R^2 and Direct, Indirect and Total effects

Dependent Variable	R^2	Independent Variables	Direct effect	Indirect effect	Total effect
(un-)certainty	0.36	TTAC	0.68	0.00	0.68*
		TTAW	0.31	0.00	0.31*
		GE	0.00	0.19	0.19*

This model explains only the 17% of the Actual Performance. However, in the case we embody the (un-)certainty parameter in the previously suggested model (in [1], shown in Fig. 4), the overall explanation of the variance in AP increases to 63.1%, but the path coefficient (direct effect) of (un-)certainty decreases to 0.13 (t=1.71) (Fig. 5)

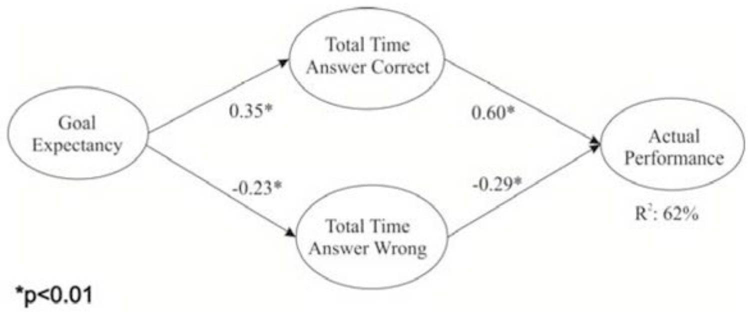

Fig. 4. Path coefficients of the research model (from [1])

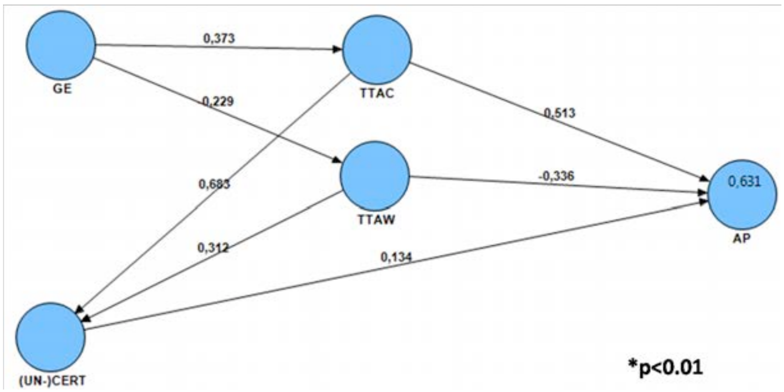

Fig. 5. Path coefficients of the research model

Furthermore, we explored the effect of the (un-)certainty parameter on students' perception of performance pre and post test. No significant effect was found in both of these cases, as shown in Table 7.

Table 7. Path coefficients for the effect of (un-)certainty on perceptions of performance

Path	Path coeff.	t value
(un-)certainty -> AP	0.41*	5.8
(un-)certainty -> PPPre	0.01*	0.12
(un-)certainty -> PPPost	0.05*	0.33

*p<0.01

This result indicates that although (un-)certainty may explain satisfactorily what students actually know, it cannot be used to determine what they believe they know neither pre-test nor post-test.

4 DISCUSSION, CONCLUSIONS AND FUTURE WORK

Both perceptions of performance and (un-)certainty are acknowledged as "subjective" factors (simple or latent), while time is an objective, actual variable. The aim of this study and its contribution was threefold: a) to explore students' perceptions of performance before and after a computer-based test, b) to investigate students' temporal behaviour – expressing (un-)certainty – when dealing with prediction of performance and c) to explore and/or identify relationships between (a) and (b).

We conducted a case study with a simplified version of the LAERS assessment environment. 96 students from Secondary Education participated in our case study. For the collected data analysis regarding our first goal we used statistical measures. Regarding our second and third goals we used SEM (in particular the PLS technique).

The results indicate that students' perceptions of performance significantly differ from their actual performance both pre-test (t=5.53, two tailed p=0.001) and post-test (t=6.56, two tailed p=0.001). This finding is in accordance to [35] and indicates that students' perceptions of performance sometimes can be inaccurate in the context of comprehension and application activities. It would be interesting to examine the respective perceptions in the context of the upper three levels of Bloom's taxonomy (analysis, synthesis and evaluation) [19] assessment activities.

Furthermore, the statistical tests failed to reveal significant differences between PPPre and PPPost (t=0.687, two tailed p=0.493). This means that students maintain their initial perceptions even after the testing procedure, which was expected to have an altering effect on their beliefs. Hacker et al. [36] suggested that when students demonstrate strong biases in their perceived performance judgement, they may not take the remedial steps necessary to improve their responses during or after an exam. This may relate to their limited self-awareness, to their satisfaction regarding their perceived goal-achievement, to their distorted perception of the test difficulty, and many more which should be further explored.

Another interesting finding was that (un-)certainty – as defined here, that is the students' cautiousness during testing in terms of time-spent on answering the quiz – explains satisfactorily the students' actual performance (shown in Table 5), but fails to explain PPPre and PPPost (shown in Table 7). These two results address the following issues: a) (un-)certainty seems to increase students' effort to answer the quiz, and consequently, the motivational effect of the (un-)certainty parameter should be further explored, and b) (un-)certainty does not seem to trigger the students' self-awareness mechanism in a conscious manner. Consequently, other temporal factors should be examined in that direction.

Students' effort and its opposite (i.e. guessing behavior) have been investigated in literature. Authors examine mostly psychometric measures, but a methodology that combines IRT and SEM could also be explored, as suggested in [37]. These studies do not follow the temporal-behavior approach, and it would be interesting to identify measure and validate "effort" as a temporal variable, and examine its correlation to cautiousness.

Furthermore, we discovered an indirect effect of goal-expectancy (GE) on (un-)certainty (see Table 6). As mentioned on section 2.3.2, GE measures students' fulfilment with their preparation before assessment, while (un-)certainty measures their cautiousness during assessment. In a sense, GE could be considered as an indicator of students' perception of preparation. If we accept that assumption, and in accordance to the results of the current study, students' perception of preparation affects their cautiousness. In addition, students' perception of performance post-test could be considered as an indicator of their satisfaction regarding their perceived achievement (Section 2.3.1). However, our results (see Table 7) did not reveal any significant correlation between cautiousness and satisfaction. This potential relation would be interesting to be further explored.

Our findings suggest that large amount of logged temporal information provide an analytic opportunity for investigation of students' behaviour before, during and after assessment. It also allows for modelling these behaviours regarding estimation of performance.

REFERENCES

[1] Papamitsiou, Z., Terzis, V. and Economides, A.A. (2014). Temporal Learning Analytics during computer based testing, (to appear) *In proceedings of the fourth International Conference on Learning Analytics and Knowledge (LAK'14)*, Indianapolis, USA.

[2] Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues, *Acta Psychologica*, 77, pp. 217–273.

[3] Adams, T.M. and Ewen, G.W. (2009). The importance of confidence in improving educational outcomes, *25th annual conference on Distance Learning and Teaching*, Madison, Wisconsin.

[4] Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning, *Educational Psychologist*, Vol 28, No. 2, pp. 117-148

[5] Bol, L. and Hacker,D.J. (2012). Calibration, In N. Seel (ed.), *Encyclopedia of the Sciences of Learning*, pp. 495-498.

[6] Chevalier, A., Gibbons, S., Hoskins, S., Snell, M. and Thorpe, A. (2008). Students' academic self-perception, CEE Discussion Papers 0090, Centre for the Economics of Education, LSE.

[7] Sit, C.H.P, Braman, O. R., Kerr, J.H. and Lindner, K.J. (2013). Motivational style and actual and perceived academic performance of secondary school students in Hong Kong, *School Psychology International*, vol. 34, No. 1, pp. 17-32.

[8] Wentzel, K.R. and Wigfield, A. (1998). Academic and Social Motivational Influences on Students' Academic Performance, *Educational Psychology Review*, Vol. 10, No. 2, pp. 155-175.

[9] Abdous, M., He, W., and 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.

[10] Huang, C-J., Chu, S-S., and Guan, C-T. (2007). Implementation and performance evaluation of parameter improvement mechanisms for intelligent e-learning systems, *Computers & Education*, Vol. 49, No. 3, pp. 597-614.

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

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

[13] Moridis, C. N. and Economides, A. A. (2009). Prediction of student's mood during an online test using formula-based and neural network-based method, *Computers & Education*, Vol. 53, No. 3, pp. 644-652.

[14] Romero-Zaldivar, V-A.,Pardo, A., Burgos, D. and Kloos, C. D. (2012). Monitoring student progress using virtual appliances: A case study, *Computers & Education,* Vol. 58, No. 4, pp. 1058-1067.

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

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

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

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

[19] Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. New York: Longmans, Green.

[20] Gross, S.R, Holtz, R and Miller, N. (1995). Attitude certainty, In RE Petty & J A Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 215-245). Hillsdale, NJ: Lawrence Erlbaum Associates.

[21] Peterson, D. K., and Pitz, G. F. (1988). Confidence, uncertainty, and the use of information, *Journal of Experimental Psychology: Learning, Memory, & Cognition*, Vol. 14, pp. 85-92.

[22] Sniezek, J. A. (1992). Groups under uncertainty: an examination of confidence in group decision making, *Organizational Behavior and Human Decision Processes*, 52, pp. 124-155.

[23] Petty, R.E., Briñol, P., Tormala, Z.L. and Wegener, D.T. (2007). The role of meta-cognition in social judgment, In ET Higgins & AW Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (2nd ed. pp. 254-284). New York: Guilford Press.

[24] Visser, P.S., Bizer, G.Y. and Krosnick, J.A .(2006). Exploring the latent structure of strength related attitude attributes, In MP Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol.38. pp.1-67). New York: Academic Press.

[25] Terzis, V., and Economides, A. A. (2011). The acceptance and use of computer based assessment, *Computers & Education,* Vol. 56, No. 4, pp1032–1044.

[26] Fornell C., and 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.

[27] 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.

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

[29] Wixon, B.H., and Watson, H.J. (2001). An empirical investigation of the factors affecting data warehousing success, *MIS Quarterly*, Vol. 25, Iss. 1, pp. 17-41.

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

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

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

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

[34] Fornell, C. and Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research,* Vol. 18, No. 1, pp. 39-50.

[35] Clauss, J. and Geedey, K. (2010), Knowledge surveys: students' ability to self-assess, *Journal of the Scholarship of Teaching and Learning*, Vol. 10 No. 2, pp. 14-24.

[36] Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of memory and metacognition,* pp. 411–455.

[37] Ferrando, P. J., Anguiano-Carrasco, C. and Demestre, J. (2013). Combining IRT and SEM: A Hybrid Model for Fitting Responses and Response Certainties, *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 20, No. 2, pp. 208-225.