DETERMINING LEARNERS' BEHAVIORAL PATTERNS IN A TECHNOLOGY AND ANALYTICS ENHANCED ASSESSMENT ENVIRONMENT

Muhittin Şahin[1] and Dirk Ifenthaler[1,2]
[1]*University of Mannheim, Mannheim, Germany*
[2]*Curtin University, Australia*

ABSTRACT

Within digitally-supported learning environments, learners need to observe themselves so that they can reflect on their strengths and weaknesses and take a step toward autonomous learning. Within the scope of this research, a technology and analytics enhanced assessment environment in which students can assess themselves was implemented and tested. In order to determine N = 108 students' ability to use the assessment environment, behavioral patterns were determined based on their individual characteristics, system interactions, lesson performances, achievement motivation, confidence, and study interest. Findings indicate different system usage and behavioral engagement with the assessment system. The outcomes of this design-based research project indicate hints on how such assessment systems can be made more effective in future implementation stages.

KEYWORDS

Computer Classification Testing, Learner Behavioral Pattern, Behavioral Engagement, Optimal Scaling

1. INTRODUCTION

Assessment is a basic component of effective learning (Bransford, Brown, & Cocking, 2000). Learning and teaching processes should be assessment-centered to develop learning (Gikandi, Morrow, & Davis, 2011). At this point, web-based assessment systems provide important opportunities (Pachler et al., 2010). These systems are named in different ways such as Computer Based Assessment, Computerized Testing, and Computer-Administered Testing, and the goal of these systems carry out assessment activities via technology (Redecker, 2013). In such assessment environments, immediate feedback can be provided to learners and scoring can be done instantly. Nowadays, it is seen that especially Computer Adaptive Testing (CAT) and Computerized Classification Testing (CCT) applications are developed and presented frequently as a technology and analytics enhanced assessment system. Technology and analytics-enhanced assessment systems are required to include flexible, adaptable, automated reports for the management of large amounts of data, analysis of unstructured data, immediate feedback to both instructors and learners, and educational decision-making (Ifenthaler, 2022).

On the other hand, learners need to observe themselves (Boud, 2000), so that learners can realize their strengths and weaknesses (Gikandi, Morris, & Davis, 2011) to self-assess and take a step toward autonomous learning. Self-assessment plays a key role in formative assessment (Andrade & Valtcheva, 2009). Educational tests are used to develop students' understanding (Bull & McKenna, 2004). Educational testing is used to make judgments about the individuals' progress, status, or accomplishments to make decisions about learning, instruction, and educational policy (AERA, 2014). It is seen that the tests contribute to the monitoring of the students' development (Rudman, 1989).

In the scope of this research, CCT was used as a technology and analytics enhanced assessment system. The CCT aims to classify the students into two or more categories rather than determine their ability estimate (Lin & Spray, 2000). Besides this, the item pools of the classification tests need not be as large as the adaptive tests (Parshall et al., 2002). Since there are limited items and the students are classified according to their achievements, the CCT environment was presented to the students within the scope of this research. CCT has various goals in the learning process such as; a) making a judgment about the students, b) assessing the learning process, c) assessing the students and school groups, and d) making a judgment about the quality of the education (van Groen, 2012).

Interactions of learners in assessment systems constitute their assessment experiences. These experiences differ according to the individual characteristics of the learners. The motivation of learners is important for their assessment experience (Vaessen et al, 2017). On the other hand, study interest is related to the perceived support of motivation, autonomy, and competence (Müller & Louw, 2004). In addition to the above-mentioned, acceptance structures are known to affect the use of e-assessment systems (Deutsch, et al., 2012). In this context, study interest, achievement motivation, and acceptance structures (perceived usefulness and intended use) were considered as individual learner characteristics in this study.

In this context, a CCT environment was implemented in which students can assess themselves and determine whether they are masters or non-masters in a subject. To determine the students' ability to use these environments more effectively and efficiently, behavioral patterns were determined based on their individual characteristics. Learners' CCT system interactions, lesson performances, achievement motivation, confidence, study interest, and acceptance structure were examined.

2. METHOD

2.1 Participants

Participants of this design-based research study included N = 108 students from an European university in the area of economic and business education. Two-thirds (67%; N_F = 72) of the participants are female and 33% (N_M = 36) are male students. Participants' average age was 23.42 years (SD = 2.69) while they studied for an average of 5.8 semesters (SD = 1.50). Students interacted with the assessment system where they could assess themselves throughout the semester (12 weeks). Ethics consent was obtained for this research.

2.2 Computer Classification Testing Environment

The aim of Computer Classification Testing (CCT) is to group learners into two or more categories (Weiss, 1982; van Groen, 2012). Two categories (Spray & Reckase, 1994; Huebner, 2012; van Groen, Eggen, & Veldkamp, 2016) and three and more categories (Eggen & Straetmans, 2000; Nydick, Nozawa, & Zhu, 2012) CCT Classification tests are frequently used for classification and licensing (Parshall et al., 2002). CCT is an effective test method that aims to classify students into groups with the least number of items by reducing classification errors (Thompson, 2007). For this purpose, in this study CCT environment was presented to the learners that classify the learners based on the Sequential Probability Ratio Test (SPRT) algorithm. SPRT was put forward by Wald (1947) and it tries to decide which of the two simple hypotheses is more correct. If the aim of a test is to classify individuals into two or more categories instead of predicting the ability or skill levels of individuals, SPRT which selects and applies the most appropriate items based on statistical hypothesis testing can be applied as a CCT procedure (Spray & Reckase, 1996). The primary focus of CCT applications with SPRT is to determine individuals' competency levels according to a standard (Parshall et al., 2002). The SPRT algorithm is less complex, more practical, and needs less time for rendering by computers (Frick, 1990). Frick (1992) found that the SPRT algorithm can classify a student as master or non-master in an average of ten items. Since there are not too many items in the item pool, SPRT which can give the most effective result with fewer items were used in this study.

A CCT environment was provided to students, where they were able to test themselves on a specific subject whenever and wherever they wanted. Learners are classified into two categories as master and non-master by the system. The CCT environment provided tests to the learners on five subjects: a) research process, b) research design, c) statistical correlation, d) statistical differences, and e) research quality criteria. However, there were limited items available for the corresponding lecture. Accordingly, the SPRT algorithm, which can work effectively with a limited number of items, was used as the classification algorithm. In addition, random item selection was used as an item selection algorithm. After students are classified by the system, their performances based on the items are automatically presented via a dashboard chart (see Figure 1).

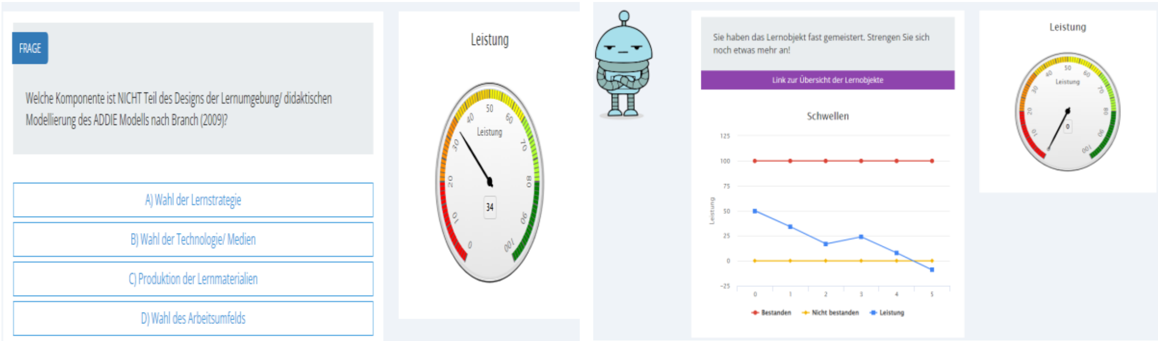

Figure 1. Screen-shot of the CCT system

2.3 Data Collection and Data Analysis

Both log data and self-report data were collected as a data collection tool. Log data consists of interaction data of learners in the CCT environment such as number of responses, number of test attempts, number of correct answers, number of incorrect answers, number of master subjects, and number of non-master subjects. On the other hand, information about learners' achievement motivation, confidence, study interest were collected as self-report data. In addition to these, end-of-term performances from the final exam were also included in the analysis.

Optimal scaling was conducted as a data analysis method. Optimal scaling is an analysis in which canonical correlation can be examined between categorical variables. The most important advantage is that more than two categorical variables can be analyzed and the results can be presented in a two-dimensional centroid graph. As the correlation among the variables increases, the points on the graph that represent them become closer, as the correlation decreases the points become distant.

3. FINDINGS

Optimal scaling analysis was employed in order to reveal the relationship between the behavior patterns of the students in the CCT environment and their psycho-educational structures. In order to perform this analysis, firstly, continuous data were converted to discrete form. For this purpose, the mean values for each structure were taken as a reference and the data were made discrete. Information on the discrete variables is presented in Table 1.

Table 1. Descriptive information about the discrete variables

Variable	Category	Frequency	Code	Measure
Achievement	Low	54 (50%)	1	Nominal
Motivation	High	54 (50%)	2	
Behavioral	Low	73 (68%)	1	Nominal
Engagement	High	35 (32%)	2	
Confidence	Low	62 (57%)	1	Nominal
	High	46 (43%)	2	
Performance	Low	48 (44%)	1	Nominal
	High	60 (56%)	2	
Study Interest	Low	58 (54%)	1	Nominal
	High	50 (46%)	2	
Use	Low	43 (40%)	1	Nominal
	High	65 (60%)	2	

As seen in Table 1, individuals above the mean were coded as High and those below the mean as Low. The interactions of the students in the CCT environment were considered behavioral engagement. In order to produce this variable, the log data of the students in the system were used. Log data consists of the number of logins, number of correct answers, number of incorrect answers, number of master subjects, number of non-master subjects, and number of attempts. With these data, principal component analysis was used and the behavioral engagement score was obtained. Other variables consist of self-report data collected from students with scales. In the second stage, it was determined which data set the variables would be included in the analysis. Variables and analysis sets are presented in Table 2.

Table 2. Datasets and variables

Set	Variables
Set 1	Achievement Motivation
	Confidence
	Study Interest
	Acceptance Structures
Set 2	Performance
Set 3	Behavioral Engagement

As can be seen in Table 2, as set 1, achievement motivation, confidence, study interest, and use; course performances of the students as set 2, and behavioral engagement as set 3 were included in the analysis. The analysis results obtained are presented in Figure 2.

Figure 2. Results of the optimal scaling analysis

As seen in Figure 2, there are 3 basic clusters. It is possible to interpret the obtained findings as follows.

- Learners who have high-level confidence have high performance and high-level AMT perceived usefulness and intended use
- Learners who have low-level confidence have low performance and low-level AMT perceived usefulness and intended use
- Learners who have low-level achievement motivation have low-level behavioral engagement and study interest. Learners' behavioral engagement in the system is negatively affected by their achievement motivation and study interest structures.

4. DISCUSSION AND CONCLUSION

Student-centered assessments aim at supporting learning and development, or increasing motivation (Andrade, Huff, & Brooke, 2012). Within the scope of this research, a technology and analytics enhanced assessment environment in which students can assess themselves was presented. The behaviors of the students in this environment and the structures associated with these behaviors were determined. For this purpose, optimal scaling analysis was employed. The findings show that learners who have low-level study interest and achievement motivation have low-level system usage or behavioral engagement. Holmes (2015) supports this finding because it revealed that there is a positive relationship between learner engagement and motivation. In this case, interventions, feedback, or recommendations are needed to increase students' motivation and study interest in order to increase their use of the system. On the other hand, it is seen that the end-of-term performance of the students who have low confidence and low system use is also low.

On the other hand, it was found that the learners who have high performance have high system use and confidence levels. It can be said that students' self-confidence affects their performance positively. In addition, students with a high level of perceived usefulness also have a high intention to use the system (Yurdugül & Bayrak, 2014). Therefore, it was expected that students with high use of behavioral engagement levels would also be high. However, within the scope of this research, no findings were reached regarding this assumption.

In the current study, the system behaviors of the learners were revealed based on psycho-educational structures. On the other hand, dispositions such as cognitive strategies, sources of motivation, and learning strategies that may affect students' system usage behaviors should also be tested. If the structures that affect the students' CCT behaviors can be determined, it is thought that important clues will be obtained on how the systems can be made more effective. In this way, behavior patterns will be revealed first and necessary improvements can be made in CCT systems.

Student-centered assessments are important for self-directed and self-regulated learning (Andrade, Huff, & Brooke, 2012). Self-assessment is given when students assess themselves and decide about the next step (Boud, 2013). In the system presented in this context, learners are presented with master or non-master status for a subject. The presentation of information on students' interactions with the system and their performance also positively affects their judgment and decision-making processes in the assessment environment. For this purpose, it is planned to integrate dashboards into the assessment system in the next stage. In this way, it is thought that students' self-assessment processes will be supported.

REFERENCES

American Educational Research Association. (2014). *Standards for educational and psychological testing.* American Educational Research Association American Psychological Association National Council on Measurement in Education.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through selfassessment. *Theory into Practice, 48*(1), 12-19.

Andrade, H., Huff, K., & Brooke, G. (2012). Assessing learning. *Education Digest, 78*(3), 46-53.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167.

Boud, D. (2013). Enhancing learning through self-assessment. Routledge.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). “How People Learn: Brain, Mind, Experience, and School”, Expanded Edition, Washington, DC: National Academy Press.

Bull, J., & McKenna, C. (2004). *Blueprint for computer-assisted assessment*. London, RoutledgeFalmer.

Deutsch, T., Hermann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment – A web-based mock examination changes attitudes. *Computers & Education, 58,* 1068-1075.

Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement, 60*(5), 713-734.

Frick, T. W. (1990). A comparison of three decision models for adapting the length of computer-based mastery tests. *Journal of Educational Computing Research, 6*(4), 479-513.

Frick, T. W. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research, 8*(2), 187-213.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333-2351.

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education, 40*(1), 1-14.

Huebner, A. (2012). Item overexposure in computerized classification tests using sequential item selection. *Practical Assessment, Research, and Evaluation, 17*(1), 12.

Ifenthaler, D. (2022). Automated Essay Scoring Systems. In *Handbook of Open, Distance and Digital Education* (pp. 1-15). Singapore: Springer Nature Singapore.

Lin, C. J., & Spray, J. (2000). *Effects of Item-Selection Criteria on Classification Testing with the Sequential Probability Ratio Test.* ACT Research Report Series.

Müller, F. H., & Louw, J. (2004). Learning environment, motivation and interest: Perspectives on self-determination theory. *South African Journal of Psychology, 34*(2), 169-190.

Nydick, S. W., Nozawa, Y., & Zhu, R. (2012, April). Accuracy and efficiency in classifying examinees using computerized adaptive tests: an application to a large scale test. In *annual meeting of the National Council on Measurement in Education, Vancouver, Canada.*

Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education, 54*, 715–721.

Parshall, C. G., Spray, J. A., Kalohn, J., & Davey, T. (2002). *Practical considerations in computer-based testing.* Springer Science & Business Media.

Redecker, C. (2013). The use of ICT for the assessment of key competences. *Joint Research Centre of the European Commission Scientific and Policy Report.*

Rudman, H. C. (1989). *Integrating testing with teaching.* ERIC Clearinghouse.

Spray, J. A., & Reckase, M. D. (1994). The Selection of Test Items for Decision Making with a Computer Adaptive Test. *Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA, April 5-7, 1994.*

Spray, J. A., & Reckase, M. D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*(4), 405-414.

Thompson, N. A. (2007). A practitioner's guide for variable-length computerized classification testing. *Practical Assessment, Research, and Evaluation, 12*(1), 1.

Vaessen, B. E., van den Beemt, A., van de Watering, G., van Meeuwen, L. W., Lemmens, L., & den Brok, P. (2017). Students' perception of frequent assessments and its relation to motivation and grades in a statistics course: a pilot study. *Assessment & Evaluation in Higher Education, 42*(6), 872-886.

van Groen, M. M. (2012). Computerized Classification Testing and Its Relationship to the Testing Goal. *Psychometrics in Practice at RCEC*, 125.

Wald, A. (1947). Sequential analysis. New York: John Wiley.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing.

Yurdugül, H., & Bayrak, F. (2014). The acceptance of web based formative assessment system for primary school students. *Journal of Educational Sciences & Practices, 13*(26), 167-186.