Discovering Students' Complex Problem Solving Strategies in Educational Assessment

Krisztina Tóth
German Institute for International Educational Research
Schloßstraße 29
60486 Frankfurt am Main
toth@dipf.de

Heiko Rölke
German Institute for International Educational Research
Schloßstraße 29
60486 Frankfurt am Main
roelke@dipf.de

Samuel Greiff
University of Luxembourg
6, rue Richard Coudenhove Kalergi
1359 Luxembourg-Kirchberg
samuel.greiff@uni.lu

Sascha Wüstenberg
University of Luxembourg
6, rue Richard Coudenhove Kalergi
1359 Luxembourg-Kirchberg
sascha.wuestenberg@uni.lu

ABSTRACT

A huge amount of log data accumulates automatically during computer-based educational assessments that can be analyzed for diagnostic or educational purposes using data mining techniques. In this paper, we describe our work of mining students' complex problem solving interactions when tackling previously unknown and dynamically changing situations.

Based on log data analyses, we discovered several problem solving strategies and examined relationships of these strategies and test outcomes. We applied clustering algorithms to discriminate between students with different levels of proficiency in problem solving. We identified four groups of students: two clusters represent successful problem solvers who differ in their level of efficiency, one group of inefficient students might need further practice to be able to solve these kinds of tasks, and finally we found a mixed-strategy group of students. Students in this last group were dynamically developing their problem solving strategy and in parallel, the ratio of correct responses increased from task to task during assessment. In sum, our findings help to advance research on cognitive processes; we support educational researchers in better understanding complex problem solving behavior and identify levels of problem solving proficiency.

Keywords

Complex problem solving, clustering, MicroDYN, test-taking behavior.

1. INTRODUCTION

Computer-based assessment allows new ways of investigating processes involved in complex problem solving (CPS) ([14]) when students solve real-life problems and the solution cannot be obtained by merely applying preexisting knowledge. To give an example: imagine you bought a new smartphone and would like to install a new application on it, but you have never used this kind of device before and do not want to read the manual. In this case, you cannot rely on previous knowledge and have to find out how the phone works by interacting with the device. After the exploration of the phone, you have a mental representation about how it works and can use your acquired knowledge to reach certain goals (e.g., to install an application). This kind of situation represents a complex problem where participants have to interact with task environments "that are dynamic (i.e., change as a function of user's intervention and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process" [2].

As the interactions between test-taker and this kind of problem situations (i.e., tasks) are essential for solving the CPS tasks, the CPS competency can only be measured in computerized environments. Due to computer-based test delivery, the test-takers' interactions with the tasks are automatically saved during the assessment. These large amounts of information pertaining to trace data constitute the basis for further analyses of participants' CPS behavior.

Based on the log data accumulated during the CPS assessment, we propose an identification of groups of students showing similar behavior in CPS assessment and identification of various CPS strategies with a clustering algorithm. Furthermore, we aim at examining relationships between the found CPS strategy and test outcomes which help us discriminate between students with different proficiency levels in problem solving.

2. RELATED WORK

CPS is a strong predictor of academic [13] and occupational achievement [3]. CPS has recently received considerable public interest, as CPS competency was tested in the Programme for International Student Assessment (PISA), a large-scale study of educational achievement assessing abilities of approximately half a million students in over 70 countries [10].

In CPS, three main processes can be distinguished: rule identification, rule knowledge acquisition and rule knowledge application [9]. In this paper, we aim at investigating CPS strategies in the rule identification phase.

In the case of data mining, empirical research has focused on investigating problem solving behavior with data mining algorithms (e.g. [1], [4], and [12]). All of these studies involved students facing other types of problem solving situations (like spreadsheets, IMMEX environment). To our knowledge no research based on data mining has yet investigated dynamic CPS strategies, as applied in this paper.

From the perspective of method, the closest study to ours was presented by [1]. It used predefined variables describing participants' problem solving behavior (like time duration, "maximum number of meters that were opened" [1]) and applied K-means to identify groups of students. Other relevant behavior analysis research often applies sequential pattern finding algorithms to search for sequences of actions (e.g. [4]), but the interpretation of clustered action sequences in the field of CPS strategy is challenging. For this reason, we propose to utilize predefined features which characterize individual CPS behavior per task to help discover students' CPS strategies.

3. METHODS

3.1 Sample

393 German students attending grades 10 to 12 participated in this study (age*:* M=17.07, SD=1.12; 60% female, 1% did not report on gender). Participation was voluntary (see [11]).

3.2 Instrument

Tasks based on the MicroDYN approach [6] were used in this study to measure CPS. MicroDYN tasks are based on linear structural equations, in which up to three input and output variables are related. Participants' tasks while dealing with MicroDYN can be best described by means of the sample task "Handball" depicted in Figure 1.

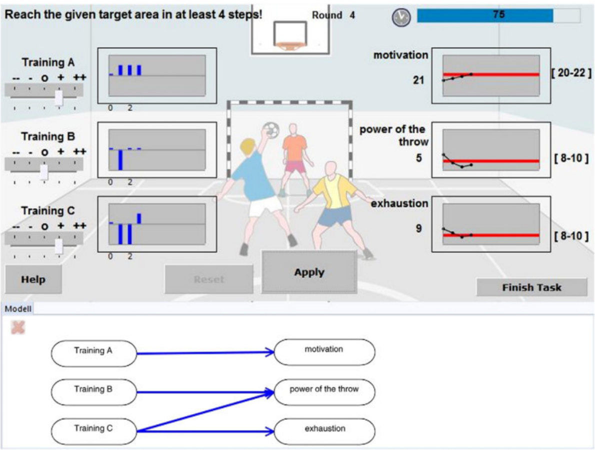

Figure 1 Sample task from the test.

In this task, different kinds of training (Training A, Training B and Training C) serve as input variables and different team characteristics (motivation, power of throw, and exhaustion) serve as output variables. Participants have to find out how the different kinds of training affect the team characteristics in order to reach a given goal state. While working on this task, participants are confronted with three different phases. First, participants freely explore the task by implementing adequate strategies to find out how input and output variables are related (rule identification phase). Therefore, participants manipulate the sliders (cf. Figure 1; below Training A, B and C) in order to change the value of the training (from – – to ++), click on "Apply" and try to retrace changes in the output variables according to their actions. Second, participants are asked to draw a causal diagram between input and output variables to visualize relations between training types and team characteristics (knowledge acquisition phase). Third, participants have to apply the retrieved knowledge in order to reach given target goals in the output variables within four steps (knowledge application phase).

3.3 Data collection

Data collection took place on several days between March and April 2011. The CPS test contained eight interactive tasks. For the present survey, three tasks were selected representing different difficulty levels. The test was administered on the school computers. In groups of maximum 16, students worked on MicroDYN for about 45 minutes.

3.4 Dataset

As outlined, log file data of the rule identification phase were analyzed in this study to identify various CPS strategies. Based on [6], one of the relevant CPS strategies by which relations between variables are identified is "vary-one-thing-at-a-time" (VOTAT), whereby only one input variable is manipulated and the others are kept constant [13]. However, VOTAT was examined as a dichotomous variable at task-level by [5], while we investigated the strategy continuous variable to get detailed information about action level activities. We integrated further action-level features (see Table 1) to offer additional information about individual activities and time information for diagnostic purposes [7]. Consequently we extracted features (see Table 1) for each student to characterize his or her individual CPS behavior. Table 1 contains the pairs of extracted features with respective interpretation.

Table 1. Extracted features and interpretations

Features	Interpretation
Number of executions	The frequency of using the apply button in the rule identification phase
Ratio of rounds with one input variable	No. of executions in which only one input variable was manipulated, divided by the total no. of executions
Ratio of zero rounds	The frequency of executions in which all sliders are set on 0 divided by the total number of executions
Ratio of repeated executions	Number of executions which were previously applied to the task, divided by the total no. of executions
Exploration time	Time given in seconds spent in the rule identification phase

3.5 Method of Analysis

The method of analysis used in this paper is X-means algorithm (with Euclidean distance) implemented in the WEKA tool [8] to detect patterns of (unlabeled) CPS behavior data. The data set contains 393 feature vectors that describe students' activities during test-taking. Each feature vector has 15 attributes, the introduced five process measures (see Table 1) built for all three complex tasks.

4. RESULTS AND DISCUSSION

4.1 Investigating CPS behavior with process measures

The CPS behavior of the students is illustrated in Table 2. It contains the average values of each feature. Data provided in Table 2 indicates that the number of executions ($t_{task1\text{-}task2}$=8.09 and $t_{task2\text{-}task3}$=4.43, p<.01), the exploration time ($t_{task1\text{-}task2}$=9.21

and $t_{task2-task3}$=7.37, p<.01) and the ratio of repeated executions ($t_{task1-task2}$=2.97 and $t_{task2-task3}$=3.24, p<.01) considerably decreased during the test-taking. In line with this finding, the ratio of rounds with one input variable increases significantly ($t_{task1-task2}$= -7.03 and $t_{task2-task3}$=-5.25, p<.01). This tendency confirms that students enhanced their CPS competency by practicing while working on the test.

Table 2. Mean value of each feature

Features	Task 1	Task 2	Task 3	*Total*
Number of executions	11.67	8.04	7.03	*8.96*
Ratio of repeated executions	.32	.28	.25	*.28*
Rounds with one input variable	.65	.75	.81	*.74*
Ratio of zero rounds	.07	.06	.06	*.07*
Exploration time	72.55	58.81	51.15	*60.83*

It is apparent that students accomplished tasks in an increasingly effective way and they more and more preferred the VOTAT strategy even though the CPS situations became more challenging as the complexity of tasks increased (the number of involved variables and the connectivity between input and output variables).

4.2 Clustering results

We identified four groups of participants which are characterized by cluster centroids (see Table 3). The members of the first subgroup (27.41 % of the whole sample) prove to be the most active problem solvers in this sample; they have the highest number of executions. Participants in Cluster 2 seem to be the most passive problem solvers as they have the lowest number of executions and they spent the least amount of time on solving the problems. The members of the third cluster are not as active as students in Cluster 1, but they take the longest to decide about the connections of input and output variables. Cluster 4 members required about the average time to solve the tasks, but used a lower number of executions than Cluster 1 and 3 members. However, the most striking difference is that the students in Cluster 4 used the lowest number of repeated executions.

To gain a deeper insight into the four cluster characteristics, we created Table 4 to represent the ratio of correct answers in knowledge acquisition on all tasks among clusters. The ratio of the correct solutions in Cluster 1 and 2 is about 90%, so these clusters represent problem solvers who correctly solved the tasks. But Cluster 3 and Cluster 4 members proved to be significantly less efficient than participants in Cluster 1 and 2.

Based on the ratio of correct responses and process measure values, we identified four groups of students, represented in the four clusters: (1) goal oriented VOTAT-strategy problem solvers, (2) less efficient VOTAT-strategy problem solvers, (3) non-VOTAT strategy users and (4) mixed-strategy users. Goal oriented problem solvers (see Cluster 2) would seem to require only a low number of executions for testing their hypothesis about the influence of factors on more dependent variables, change mostly one aspect of the system (VOTAT strategy), repeat only a few executions and need little time to solve the tasks successfully. Less efficient problem solvers (see Cluster 1) are test-takers who also mostly varied one factor while others were held constant (VOTAT strategy), but used a much higher number of executions (almost three times as many) and required more exploration time for identifying correct rules than goal oriented VOTAT-strategy users.

Table 3. Cluster centroids of the X-means clustering analysis

Features	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Ratio of students	27.41%	19.80%	37.56%	15.23%
Task 1				
No. of executions	22.38	6.72	11.55	7.47
Repeated executions	.63	.18	.34	.12
Rounds with one input variable	.72	.86	.37	.34
Ratio of zero rounds	.08	.04	.13	.05
Exploration time	81.76	60.60	82.45	75.35
Task 2				
No. of executions	13.97	4.84	8.84	5.14
Repeated executions	.56	.15	.31	.09
Rounds with one input variable	.82	.95	.38	.62
Ratio of zero rounds	.07	.02	.16	.01
Exploration time	63.13	49.18	70.63	59.72
Task 3				
No. of executions	11.46	4.43	8.28	4.33
Repeated executions	.53	.13	.27	.06
Rounds with one input variable	.89	.97	.46	.75
Ratio of zero rounds	.08	.01	.16	.01
Exploration time	52.72	42.47	64.82	51.22

Non-VOTAT strategy users (Cluster 3) varied multiple aspects at once (the ratio of correct responses is only .51 in contrast to .90 in Cluster 1 and .92 in Cluster 2) and used a significantly higher number of zero rounds than VOTAT-strategy problem solvers although these tasks did not require zero rounds and they needed the most time to explore the system.

The fourth group of students dynamically developed their CPS strategy, e.g. the ratio of one input variable in a round increases from 0.34 to 0.75. However in contrast with this the exploration time (from 75.35 to 51.22 seconds) and ratio of repeated rounds significantly decreased. The ratio of correct responses correspondingly increased from task to task (as can be seen in Table 4), so the students became more effective during the educational assessment.

In sum, the group of high-achievers using VOTAT strategy can be split into two categories: goal oriented and less efficient VOTAT-strategy problem solvers. Although these students found the connections between input and output variables, they demonstrated different CPS behaviors, which can only be detected by investigating process related test-taking data.

Table 4. Ratio of correct responses

Groups of students	Task 1	Task 2	Task 3	*Total*
Cluster 1	.86	.89	.94	*.90*
Cluster 2	.92	.90	.95	*.92*
Cluster 3	.56	.42	.54	*.51*
Cluster 4	.45	.55	.68	*.56*

Furthermore, based on the introduced process measures we made a distinction between non-VOTAT strategy users and mixed-strategy users. The mixed-strategy problem solvers' group was learning from the tasks and changed/improved their CPS strategy during the test taking. For this reason, they need more time with the same test material to improve their CPS competency. But students using non-VOTAT-strategy were not learning from the test, so they need other type of instructional intervention.

These findings help to advance the research on CPS processes; we thus support educational researchers in better understanding CPS behavior and identifying levels of CPS proficiency. In addition our study helps to detect different types of instructional interventions to improve students' individual CPS competencies.

5. FUTURE WORK

As an important next step, we need to verify our results taking additional data into account. This is a rather straight-forward step that can be split into looking at additional test results for the same set of tasks and afterwards trying to find similar results looking at other CPS tasks. In parallel, we want to examine the literature on CPS strategies. To our knowledge, the groups we have identified are not described elsewhere. Generally, VOTAT and non-VOTAT strategies are distinguished but not analyzed any further, for instance, by integrating process data.

We are currently collaborating with item and test developers working on complex problem solving. Our findings can help them improve their tasks and gain a deeper understanding of how students interact with the tasks. Finally, our results can be used to go beyond dichotomous grading of CPS items. We can try to help students following an inefficient path and use certain interventions to put them back on track.

6. REFERENCES

[1] Angeli, C. and Valanides, N. 2013. Using educational data mining methods to assess field-dependent and field-independent learners' complex problem solving. *Educational Technology Research and Development*. 61, 3 (June 2013), 521-548. DOI: 10.1007/s11423-013-9298-1

[2] Buchner, A. 1995. Basic topics and approaches to the study of complex problem solving. In Frensch P. A. and Funke, J. (Ed.) *Complex problem solving: The European perspective* (27-63). Hillsdale, NJ: Erlbaum.

[3] Danner, D., Hagemann, D., Schankin, A., Hager, M., and Funke, J. 2011. Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*. 39, 5 (Sept.-Oct. 2011), 323–334. http://dx.doi.org/10.1016/j.intell.2011.06.004

[4] Fern, X., Komireddy, C. Grigoreanu, V., and Burnett, M. 2010. Mining Problem-Solving Strategies from HCI Data. *ACM Transactions on Computer-Human Interaction*. 17, 1, Article 3 (March 2010), 1-22. doi:10.1145/1721831.1721834

[5] Greiff, S., Wüstenberg, S., and Funke, J. 2012. Dynamic problem solving: A new measurement perspective. *Applied Psychological Measurement*. 36, 3 (May 2012), 189–213. doi: 10.1177/0146621612439620

[6] Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F. and Funke, J. 2013. Computer-based assessment of Complex Problem Solving: concept, implementation, and application. *Educational Technology Research and Development*. 61, 3 (June 2013), 407-421. doi: 10.1007/s11423-013-9301-x

[7] Gvozdenko, E. and Chambers, D. 2007. Beyond test accuracy: Benefits of measuring response time in computerised testing. *Australasian Journal of Educational Technology*. 23, 4, 542-558.

[8] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, R. and Witten, I.H. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl*. 11, 1 (November 2009), 10-18. DOI: http://doi.acm.org/10.1145/1656274.1656278

[9] Kröner, S., Plass, J. L., and Leutner, D. 2005. Intelligence assessment with computer simulations. *Intelligence*. 33, 4, (July-August 2005) 347–368. DOI:http://dx.doi.org/10.1016/j.intell.2005.03.002

[10] OECD. 2010. *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices* (Volume III). PISA, OECD Publishing. DOI: http://dx.doi.org/10.1787/9789264083943-en

[11] Schweizer, F., Wüstenberg, S., and Greiff, S. 2013. Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*. 24 (April 2013), 42–52. DOI: http://dx.doi.org/10.1016/j.lindif.2012.12.011

[12] Vendlinski, T. and Stevens, R. 2002. Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment*, 1,3, Available from http://www.jtla.org

[13] Wüstenberg, S., Greiff, S., and Funke, J. 2012. Complex problem solving—More than reasoning? *Intelligence*. 40, 1 (Jan.-Febr. 2012), 1–14. doi: 10.1016/j.intell.2011.11.00

[14] Zoanetti, N. 2010. Interactive computer based assessment tasks: How problem-solving process data can inform instruction. *Australasian Journal of Educational Technology*. 26, 5, 585-606.