UNIVERSITY OF SZEGED

DOCTORAL SCHOOL OF EDUCATION

INFORMATION AND COMMUNICATION TECHNOLOGIES IN EDUCATION

SALEH AHMAD ALRABABAH

PHD DISSERTATION

MEASURING COMPLEX PROBLEM SOLVING IN JORDANIAN HIGHER EDUCATION: FEASIBILITY, CONSTRUCT VALIDITY AND LOGFILE-BASED BEHAVIOURAL PATTERN ANALYSES

SUPERVISOR:

GYÖNGYVÉR MOLNÁR DSC

PROFESSOR OF EDUCATION

SZEGED, HUNGARY

2022

1. The context and structure of the dissertation

Technology has been used in assessment since the beginning of using computers in various domains and education. It has provided many advantages, possibilities, and challenges in the field of educational assessment (Pásztor-Kovács et al., 2021). By means of technology, teachers and educational administrations could have developed new policies which fit on a higher level to the expectations of the 21st century, e. g. measuring 21st-century skills even in international large-scale assessments (see e.g. Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) creative or collaborative problem-solving module; see e.g. OECD, 2014a; Griffin et al., 2012).

The use of technology in the assessment leads to improve the possibilities, the efficacy, the validity, and the quality of assessment and offers numerous advantages over traditional assessments (Alrababah & Molnár, 2021), such as automatic item development, automatic scoring (Becker, 2004; Csapó et al., 2014; Dikli, 2006; Mitchell et al., 2002; Valenti et al., 2003), and reducing costs (Bennett, 2003; Christakoudis et al., 2011; Wise & Plake, 1990). In educational assessment, it provides the basis for innovations, e.g., measuring new constructs, using new item types (Dörner & Funke, 2017). Information and communication technologies, especially computers, had an immense impact on the development of educational assessment from quantitative and qualitative points of view. New science has emerged in the field of assessment, which focuses not only on the analyses of the actual answer and achievement data, but more deeply on the analyses of the contextual data collected during the data collection beyond the students' actual answers. Log file analyses and educational data mining have become state-of-the-art educational assessment analyses attracting increasing research interest. They make it possible to answer research questions that could not be answered through traditional assessment techniques (Molnár & Csapó, 2018).

The growing field of educational data mining uses data mining techniques and methods for searching for different patterns in the recorded, basically unstructured contextual data for analyzing and extracting hidden information about students' actions or test-taking behaviour for a more deeply and better understanding of the examined phenomenon. Educational Data Mining (EDM) has become a significant aspect of analysis, the basis of further developments in educational research and practice (Dahiya, 2018).

On the one hand, logfile analyses (e.g., time-on-task, number of clicking, navigation within the test), based on structured data files can also provide information that is not available with traditional assessment techniques and contribute to a better understanding of the examined phenomenon. By means of logfile analyses, we can get data about students' test-taking behaviour, e.g. applied exploration strategies while solving complex problems (see e.g. Molnár & Csapó, 2018), or we can analyze the relations of time-on-task and achievement data. Logfile analysis also offers many challenges (e.g. how to make sense of the amount of data extracted) and possibilities (e.g. cover unhidden pattern of the educational phenomenon under examination) in the field of educational assessment (Stadler et al., 2020).

Problem-solving is one of the most often assessed reasoning skills in large-scale educational assessment projects. It is considered one of the most essential skills in the 21st century (Krieger, et al., 2021). Its assessment provides a reasonable basis for introducing how educational assessment techniques developed from traditional paper-and-pencil to computer-based assessment and how the type of research questions and used problem types varied by the changing possibilities in the field of assessment (Wu & Molnár, 2021).

In the present research project, we explore the feasibility and the potential for using computer-based assessment for assessing 21st century skills in Jordan. More specifically, we decided to assess students' 21st century skills during their higher education studies using most of the advantages of computer-based assessment. Complex Problem Solving (CPS) proved to be a good candidate for such a role. Because the test of CPS contains tasks including multimedia elements, requiring interaction (not only clicking on a radio button or entering a text in a textbox) of the test taker with the problem scenarios. It offers great possibilities for monitoring students' test-taking behaviour over time on task or the number of clicks via logged data. CPS as a construct involves knowledge acquisition (KAC) and knowledge application (KAP), which are basic learning elements. CPS allows us to investigate how knowledge is acquired in a new problem situation (KAC) and then applied to actually solve a problem (KAP) in an uncertain situation, which is independent of domain-specific content (Greiff et al., 2013). CPS is, by its nature, an important educational outcome in the twenty-first century (Krieger et al., 2021). Understanding how students acquire knowledge and then applying it has become essential because it highly predicts educational achievement (Schweizer et al., 2013).

CPS has been widely assessed in large-scale international assessments (see OECD, 2014b). However, not all of the countries which participated in the 2012 PISA cycle took part in assessing problem-solving. Only a few countries from the Middle East chose it as an international option. Jordan, the country under investigation, did not. As a result, the current study is likely to be the first to report Jordanian students' CPS skills. Despite the extensive usage of CPS in international samples, little attention has been paid to analyzing its measurement invariance across cultures and nations.

Structure of the dissertation

The theoretical part of the dissertation (first and second papers) investigates the developmental trends in technology-based assessment in an educational context and highlights how technology-based assessment has reshaped the purpose of educational assessment and the way we think about it. Developments in technology-based assessment stretch back three decades. Around the turn of the millennium, studies centred on computer-based and paper-and-pencil test comparability to ascertain the effect of delivery medium on students' test achievement. A systematic review of media studies was conducted to detect these effects. We present the developmental trends in EDM techniques and logfile analysis in the educational context and their contribution to understanding the contextual data collected beyond the particular response data. We conduct a comparison analysis based on the Scopus database to show the developmental trends by year and domain. Then we shed light on measuring complex problem-solving in the educational context and its methods with different approaches. Finally, the applications of computer-generated logfile analyses in the domain of complex problem solving were investigated. The theoretical studies contain of two journal articles:

- Alrababah, S. A. & Molnár, G. (2021). The Evolution of Technology-based Assessment: Past, Present, and Future. *International Journal of Learning Technology*, 16(2), 134–157.
- Alrababah, S. A. & Molnár, G. (2021). Analysing Contextual Data in Educational Context: Educational Data Mining and Logfile Analyses. *Journal of Critical Reviews*, 8(1), 261–273.

Despite the importance of Complex problem-solving (CPS), we have no knowledge of its measurability, development, or comparability in Arab countries, with a short history of computer-based assessment. We fill this niche and beyond monitoring the applicability of third-generation innovative tests in a Jordanian higher educational context, we run international research to understand the behavioural differences in students′ test-taking and problem-solving behaviour in case of European students (Hungarian) and Arab students (Jordanian). The results provide important insights into cross-cultural differences in test-taking behaviour and hidden behavioural patterns of students coming from Arab and European countries as they solve computer-based complex problems and contribute to an understanding of how students from different educational contexts behave while solving tests, especially, technology-based complex problems.

Papers 3-4 introduce the empirical studies conducted within the confines of the PhD research. In all data collection, CPS was measured using the MicroDYN approach (Greiff & Funke, 2017) and the online eDia platform was used to administer the test (Molnár & Csapó, 2019). Paper 4 presents the main research questions, methods and results of the pilot study. The main aim of the pilot study was to test the applicability of technology-based assessment, especially the feasibility of using innovative third-generation tests in the Jordanian higher education context, where the use of technology in assessment has less attention; and validating a third-generation online test of complex problem-solving in higher education. We also investigated students' test-taking and problem-solving behaviours while working on complex problems in a digital environment using both directly collected answer data and logfile analyses. As a result, this study investigated the role of strategic exploration, various problem solving, and test taking behaviour in CPS success by using log file data to visualize and quantify Arabic students' problem solving behaviour in six CPS problems of varying difficulty and characteristics. The results of this study have been submitted for publication to SAGE Open:

Alrababah, S. A., Wu, H., & Molnár, G. (2022). Measuring Complex Problem-Solving in Jordan: Feasibility, Construct Validity and Behaviour Pattern Analyses. SAGE Open. (Submitted).

The results of the cross-cultural comparison study are introduced in paper 4. This study analyzes behavioural and overall performance data in CPS from two different countries with very different cultures: Jordan and Hungary. First, we monitored measurement invariance of CPS (i.e., MicroDYN) across Jordanian and Hungarian context. Then, in three steps, we examined the nature of the developmental differences. First, we used the traditional scoring method to focus on students' actual answer data. Second, we gained insight into what high- and low-achieving students did during the problem-solving process. Specifically, how motivated they were, as seen by how much effort they had shown during the test administration (number of clicks) and how much time they had spent on the problems. Third, we discovered different problem-solving profiles in both countries using logfiles and a behaviour pattern-finding algorithm. We compared students' behavioural features based on their class profiles and final scores. The results of the cross-national comparison study have been published in form of a journal article:

Molnár, G., Alrababah, S. A., & Greiff, S. (2022). How We Explore, Interpret, and Solve Complex Problems: A Cross-National Study of Problem-Solving Processes. Heliyon, 8 (e08775)

Finally, the sixth part of the dissertation consists of the conclusions derived from the discussions of the findings of the studies. It also includes the recommendations and the suggestions for future research and reveals the limitations of the studies.

3. Methods

4.1 Participants

For the pilot study, the participants were undergraduate students (Mean_age= 21.50, SD_age=3.03, N=195) from two Jordanian universities having 15 and 13 faculties. Students from two faculties took part in the assessment: Arts and Sciences.

The participants in the Jordanian sample for the second data collection were studying in different years at two large Jordanian universities the sample consisted of 457 students (mean age=20.6; SD=3.11), with 53.4% of them being female. Students’ participation was voluntary; as an incentive, they earned credit for successful completion of the test.

Participants in the Hungarian for the third empirical study were commencing their studies at one of the largest and highest-ranked Hungarian universities. A total of 1844 students, that is, 44.8% of the target population, participated in the study (mean age=19.9; SD=1.82), with 59.8% of them being female. Students’ participation was voluntary; as an incentive, they earned one credit for successful completion of the tests.

4.2 Instrument

A complex problem-solving test based on the MicroDYN approach was administered in both countries. The tests consisted of the same complex problems (ten problems) with increasing item complexity (number of input and output variables and number of relations) and fictitious cover stories. At the beginning of the test, participants were provided with the same instructions on engaging with the user interface, including the same warm-up task. MicroDYN is designed to allow students to acquire the general exploring skill through problem-solving with a limited number of variables and relations, while in most cases nothing changes in the problem scenario if the participant has not changed any variables. Thus, the test is designed so that students can learn during the test-taking process as previous problems and previous problem-solving processes can influence subsequent problem-solving in the MicroDYN task. Because of these special features of MicroDYN problems and tests, the learning process can be explored and quantified, thus providing the possibility to measure the learning potential of the students occurring in the problems and during the test-taking procedure.

From the perspective of the traditional psychometric approach, each problem consisted of two phases: knowledge acquisition (first phase) and knowledge application (second phase), which were scored separately. Consequently, each problem consisted of two scoreable items.

In the first phase of the problem-solving process, the free exploration phase, the relations between the input and output variables needed to be explored by interacting with the problem environment. During this interaction process, students were expected to manipulate the values of the input variables (Greiff & Funke, 2010) as many times as they liked within 180 seconds and to identify the resultant changes in the output variables (direct effects) to acquire new knowledge (Fischer et al., 2012). The test contained tasks where output variables could have changed not only as an effect of manipulation of the input variables but also spontaneously, with internal dynamics (eigendynamic; Greiff et al., 2013). Independent of the type of effects and relations, it was possible to detect the structure of the problems with an adequate problem-solving strategy (Greiff et al., 2012) and with an adequate, systematic manipulation strategy. To do this, test-takers were expected to click on a button with a + or – sign or by using a slider linked to the respective input variable (See Figure 1) and press the Application button, which made it possible to test the effect of the set values of the input variables on the output variables, which was defined as a trial. The effect in terms of the changes in the values of the output variables was presented on a graph next to each output variable, similarly to the history of the earlier settings of the input variables within the same scenario, which was also presented on a graph next to each input variable. According to the user interface settings, within the same phase of each problem, the input values remained at the same level until the Reset button was pressed or they were changed manually. The Reset button set the system back to its original status, that is, the values of the input and output variables were reset to zero, and the history of the earlier settings and effects disappeared from the graphs. In the present paper, we have labeled and analyzed these strategies using the log data collected during the exploration phase of the problem-solving process. During this 180 second in the first phase of the problem-solving process, they were expected to draw the relations they noticed in the form of arrows between the variables presented on the concept map under the MicroDYN scenario on screen. This first phase of the problem-solving process, including the free exploration and the model building process, is often called the knowledge acquisition phase (see Greiff et al., 2013).

Figure 1. Screenshot of the MicroDYN task "Game Night." See the original version of the task in Greif et al. (2011). The controllers of the input variables range from "- -" (value=-2) to "++" (value=+2). They are presented on the left side of the problem environment in the Hungarian-language version and on the right side in the Arabic one. The model is shown at the bottom of the figure.

In the second part of each of the problems, in what is called the knowledge application phase (Greiff et al., 2013), students were expected to reach the given target values of the output variables within a given time frame (90 seconds), at most in four clicks of the Application button. In this phase the right concept map was presented to the students on screen to make the different parts of the problem-solving process as independent as possible. Finally, students were able to navigate between the different phases within the same MicroDYN scenario and between the different MicroDYN scenarios using the Next button (there was no Back button available on the test).

4.3 Procedure

The tests were administered online via the eDia assessment platform (Molnár & Csapó, 2019) for both data collections process. The data collection lasted 45 minutes at the university's computer labs for the pilot study.

Because of the COVID-19 situation, the second data collection could not be administered in a monitored environment in the Jordanian university buildings. Students received the password and link to the complex problem-solving test and were asked to complete the test at home. After entering the eDia system, students had 60 min to solve the problems and complete the related questionnaire. They received immediate feedback on their average achievement after test completion.

The Hungarian assessment was carried out in a large computer room at the university learning and information center using several security protocols due to COVID-19 (e.g., every other computer was switched off, use of face masks and hand sanitizer was compulsory, and all the key-boards and mice were disinfected during the breaks). The assessment was carried out in the first four weeks of the semester when the university was engaged in hybrid education. The tests and questionnaire were administered using the eDia online platform as was the case in the Jordanian second data collection. The testing time was limited; students had 60 min to complete the test and the related questionnaire. They received immediate feedback on their average achievement after test completion and detailed feedback with normative comparative data on their performance a week later.

As an achievement indicator, we applied the traditional scoring for both CPS phases (see e.g.; Molnár & Csapó, 2018): if the relations presented in the concept map between the input and output variables matched the underlying problem structure, students obtained a score of 1. Otherwise, the response was assigned 0 points (for the first phase). Further, if the problem-solver managed to reach the target values of the output variable by changing the values of the input variables within four trials and 90 seconds, the second phase of the problem-solving process earned 1 point and 0 points otherwise. Applying the traditional scoring, we generated databases for the analyses.

Beyond the traditional scoring, students' activity during the problem-solving process was logged and coded based on Molnár and Csapó's (2018) mathematical model and labelling system, which had been developed based on the effectiveness of the strategy usage. Every trial was formatted and labelled in the system; the operation(s) carried out by the students throughout the trial was/were summarized as one data entity in the logfile database. During this labelling process, students' problem-solving behaviour was defined in each problem environment by the evaluation of all trials within the same problem. If the problem-solving behaviour matched meaningful regularities, it was labelled as a strategy. Three categories were defined within the problem-solving strategies observed: (a) no VOTAT at all – which earned a score of 0 points; (b) partial VOTAT, when VOTAT was employed for some but not all the input variables – which was assigned a score of 1 point; and (c) full VOTAT, when VOTAT was used for all the input variables – which garnered a score of 2 points.

4. Results

The internal consistency of the CPS tests was high (Cronbach's alpha=.87), and subtest level reliabilities proved to be also high or acceptable: .84 in the KAC and .69 in the KAP phase, respectively. The test proved to be difficult for the students (M=25.9%; SD=19%), whose achievement was significantly higher in the KAC phase (M=36.5%; SD=27%) than in the KAP phase (M=15.3%; SD=15.3%).

Table 1. shows the test and subtest-level correlations. The correlation between measures of KAC and KAP is strong and significant (r=.675), however, indicating that KAC and KAP are measuring something different.

Table 1.

Test and sub-test level correlations of CPS

	KAC	KAP
KAC	1.00	
KAP	.675**	1.00
PS	.958**	.857**

Note:**: p<.01 level significant

Confirmatory factor analyses indicated a good fit (see Table 2). Within the 2-dimensional model, both measures of KAC and KAP were significantly correlated on a latent level (r=.865, p<.001; manifest correlation: r=.675, p<.001). We also tested a 1-dimensional model with all indicators combined under one general factor. The fit indices decreased considerably.

In order to compare models, we carried out a special $\chi2$-difference test in Mplus (Muthén & Muthén, 2010), which showed that the 2-dimensional model fitted significantly better than the 1-dimensional model ($\chi2$=32.645; df=1; p<.001). In summary, the 2-dimensional model fitted better, thus, the processes KAC and KAP were empirically distinguished.

Table .2.

Goodness of fit indices for testing dimensionality of CPS

Model	$\chi2$	*df*	*p*	CFI	TLI	RMSEA
2-dimensional	379.093	169	.001	.965	.961	.054
1-dimensional	406.021	170	.001	.961	.956	.057

Note. df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker Lewis Index;

RMSEA = Root Mean Square Error of Approximation; χ^2 and df are estimated by WLSMV.

The result from the international comparison study revealed that there were large differences found in students' test-taking behavior as regards time-on-task (see Table 3). On average, the Jordanian students spent 36 seconds exploring the problem, while the Hungarian students spent more time on exploration (56 sec). On the one hand, the differences become smaller parallel to the increasing complexity of the tasks; on the other hand, they become large again when problems with internal dynamics appeared on the test. This phenomenon was caused mainly by the Hungarian students, who spent ever less time on problem exploration.

The Jordanian students' test-taking behavior was more stable over time and across different levels of problem complexity. However, there was a backward but weaker tendency identified compared to the Hungarian sample. The Jordanian students spent increasingly more time with more trials – but significantly less than their Hungarian peers – in the exploration phase of the problem-solving process as the problems became ever more complex.

Table 3.

Cross-national differences in students' test-taking behavior: time-on-task

Complexity of problem	Jordanian			Hungarian			t	p	d
	Low achievement	High achievement	Mean	Low achievement	High achievement	Mean			
Time-on-task									
2-2 (2)	49.5	26.2	33.8	74.9	59.0	63.1	14.0	<.001	-0.67
3-3 (3 or 4)	38.5	37.0	37.6	55.9	47.2	49.2	6.0	<.001	-0.31
3-3 (2 or 3+1)	35.2	39.8	35.5	56.0	70.9	60.1	13.6	<.001	-0.76
Sum	39.4	35.9	36.0	59.7	59.1	56.4	18.4	<.001	-.57

We investigated latent class analyses in both samples among the behavior patterns in the log data. They were scored according to the level of optimal exploration strategy use: 2: fully isolated variation strategy; 1: partially isolated variation strategy; 0: no isolated variation at all.

Four latent classes were distinguished in the Jordanian sample (as well as in the Hungarian sample). The classes were interpreted as follows based on their profiles: (1) non-performing explorers, (2) non-persistent explorers, (3) restarting explorers with a learning effect, and (4) almost proficient explorers.

Non-performing explorers (40% of the Jordanian students) employed no fully or partially isolated strategy at all. Non-persistent explorers proved to be intermediate explorers on the easiest problems but low explorers on the complex ones (6.6% of the Jordanian students), having employed the partially isolated variation strategy less and less parallel to the increasing level of complexity of the CPS problems. Restarting explorers with a learning effect (15.3% of the Jordanian students) were able to learn between problems of similar complexity (similar number of input and output variables and number and type of connections), but the probability of applying a partially or fully isolated strategy dropped again as the complexity of the problems grew. Almost proficient explorers (38.4% of the Jordanian students) used the isolated variation strategy with 80% probability on problems with only direct effects. Then, after a rapid learning process, they managed to continue this exploration behavior even with the CPS problems with internal dynamics (see Figure .2).

The following four latent classes were distinguished in the Hungarian sample, albeit somewhat different ones as compared to the Jordanian sample: (1) non-performing explorers, (2) restarting slow learners, (3) rapid learners, and (4) proficient explorers (see Figure 3). Non-performing explorers (7.4% of the Hungarian students) did not use any isolated or partially isolated variation at all throughout the tasks. Restarting slow learners (3.2% of the Hungarian students) were among the intermediate-performing explorers who only rarely employed a fully or partially isolated variation strategy with a very slow learning effect. Rapid learners (7% of the Hungarian students) were basically low performers with regard to the efficacy of the exploration strategy they used on the easiest problems, but they become proficient explorers as a result of rapid learning, with achievement on the complex ones that equaled that of the top performers. Proficient explorers (82.4% of the Hungarian students) used the isolated variation strategy with high probability on all the proposed CPS problems.

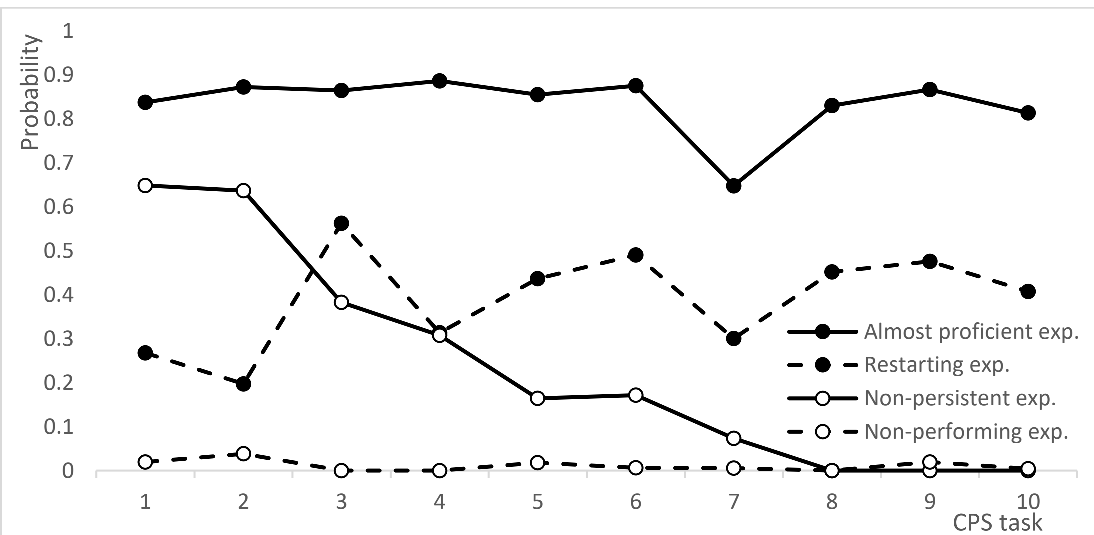

Figure 2. Four qualitatively different class profiles in the Jordanian sample

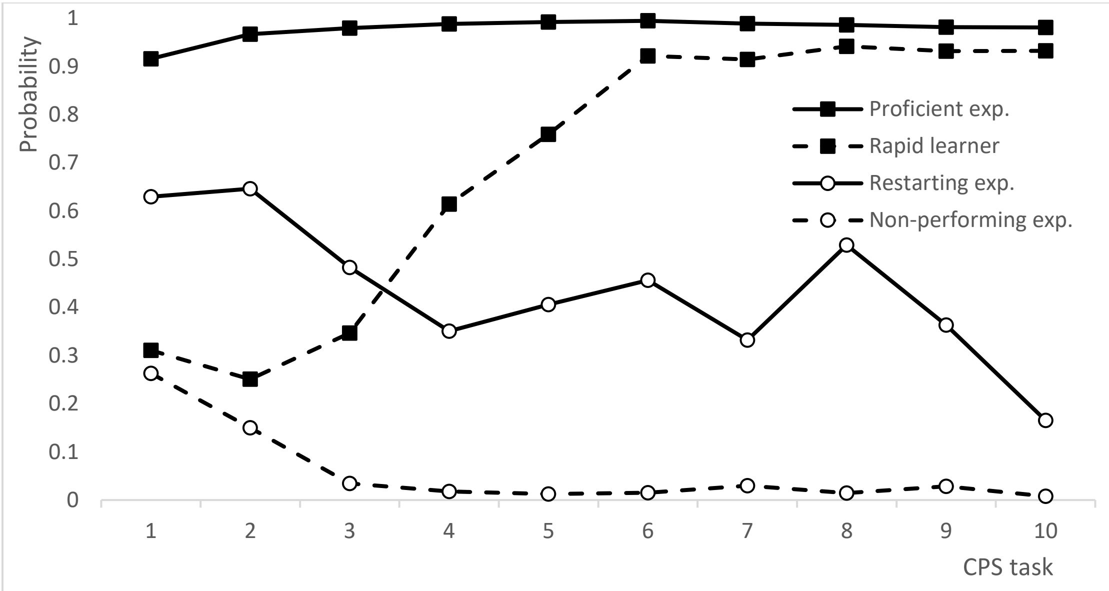

Figure 3. Four qualitatively different class profiles in the Hungarian sample

5. The discussion and conclusion derived from the results

According to the results the CPS assessments had a high level of internal consistency – similarly to the Europen results –, but the interactive problems proved to be generally hard for the Jordanian participants. We could conclude from the descriptive results that computer-based assessment and the use of innovative online tests are feasible and valid in Jordan in the higher education environment.

Logfile-based analyses extended the scope of previous research results connected to CPS, particularly in the Arabic context. We monitored and identified the way students understand interactive problems, especially minimal complex systems and causal relationships within the problems. Despite the fact that most of the Jordanian university students showed systematic strategies were unable to solve the problem and on the contrary several students managed to solve the problem without applying an effective problem solving strategy. Thus, solving an interactive problem does not necessarily require the application of a strategy that gives the problem solver sufficient information about the problem environment to achieve the correct solution, and the application of a right problem solving strategy does not always result in high problem solving achievement. This confirms de Jong and van Joolingen (1998) research results, who claim that learners often have trouble understanding data. Generally, these results are in line with previous research results (e.g., Greiff et al., 2015; Molnár & Csapó, 2018; Vollmeyer et al., 1996).

There was a significant correlation between KAC and KAP on both the manifest and latent levels. The KAC and KAP processes were empirically distinguished and confirmed by the international research results (e.g. Funke, 2001; Wüstenberg, Greiff, & Funke, 2012). More specifically, previous studies have found that KAC and KAP correlations range from weak to strong relationships between the two phases. The wide range of correlation indexes is associated with the use of multiple CPS assessments with varied approaches to measuring KAC and KAP.

To interpret and understand Jordanian research results more deeply and detect the cross-national aspects of CPS, we organized an international assessment in Hungary and Jordan. Based on the results of the cross-national large scale study, we can conclude that complex problem-solving can be measured in the Hungarian and Jordanian educational contexts validly, reliably, and equivalently. The results revealed the different behavior patterns of Hungarian and Jordanian undergrads, providing valuable insights into the international validity of CPS assessments and increasing our understanding beyond what we can learn from traditional CPS performance indicators. Even students that are socialized in the same school environment can think differently and achieve the same results with the same aims via different routes.

We investigated the way Jordanian and Hungarian students interpret CPS problems. The results showed measurement invariance of CPS across Jordanian and Hungarian undergraduates, that is, students independent of their culture interpreted CPS problems the same way. Not even the language-based conceptual representational differences and the differences in frequency usage of computer-based assessments impacted measurement invariance of CPS.

Developmental differences have been found in CPS skills between Jordanian and Hungarian university students in favor of the Hungarian students. This is consistent with previous research findings indicating that students from different educational and cultural backgrounds can perform differently in CPS environment (see Greiff, Wüstenberg, & Avvisati, 2015; OECD, 2014a; Wu & Molnár, 2021; Wüstenberg et al., 2014). The development of CPS skills is not universal. In the dissertation presented comparison analysis, we used Hungarian CPS data as a benchmark indicator. Additional studies using representative samples from both nations are required to validate these findings.

Beyond differences, there were also similarities detectable in Jordanian and Hungarian students' CPS behaviour. If Jordanian students' achievement dropped, Hungarian students' mean performance dropped too; but there was a major difference in their starting values, resulting in significant achievement differences in both phases across all complexity levels. The reasons for these differences in achievement could be due cultural and educational differences and probably their prior computer experience in academic environment. In Western countries, the use of computers in education has long been addressed and belongs to key areas supporting learning.

After we confirmed that CPS can be measured cross-nationally in the same way and the level of CPS skills of Hungarian and Jordanian students (in this sample) differ, we wanted to expand our understanding of these differences and analyse their test-taking behavior. There were significant differences in using a theoretically effective exploration strategy in both samples based on the results of the logfile analyses. In total, 93% of Hungarian university students used a theoretically effective strategy, compared to 44% of Jordanian university students. This supports our earlier discussion regarding the less complex problems in the test that the majority of Hungarian students belong to the expert problem solvers. There were also differences in the percentages of using a theoretically effective strategy and having high CPS performance. It was 60.6% on average in the Hungarian sample and 44.4% among the Jordanian students. That is, there were large differences between the two samples not only in the efficacy of their interpretation of the extracted information, but also in the suitability of the exploration strategy they employed, resulting in large significant differences in their final CPS performance.

We identified other significant behavioral differences between Hungarian and Jordanian students, which differences were becoming smaller as the task complexity increased. During the test-taking process, Hungarian students spent less and less time attempting fewer and fewer trials despite the fact that the tasks were becoming more complex in comparison to Jordanian students, who spent nearly the same time and used nearly the same number of trials throughout the test. This tendency indicated that Hungarian students became increasingly aware of their effective exploration behavior and required fewer trial-and-error moves and trials.

To characterize students' exploration strategies in a CPS context, we employed latent class analyses based on the level of the optimal exploration strategy. Four latent classes have been identified in both samples. In both samples, the classes of non-performing explorers and restarting slow learners were nearly identical, indicating differences in the behaviors of Jordanian and Hungarian students. There was a rapid learner in the Hungarian sample, which did not exist in the

Jordanian sample. Rapid learners revealed a noticeable learning curve when working on the CPS tasks. By the sixth problem having no eigendynamic on the test, they had reached the same level as proficient explorers in terms of exploration behavior. They have the ability to quickly and flexibly adapt to the expectations of a given situation (see Greiff et al., 2018). Instead of rapid learners, in the Jordanian sample we could identify a class of non-persistent explorers. These students were able to apply the partial variation strategy on the easiest problems, but they couldn't apply it to the more difficult ones.

6. References

Alrababah, S., & Molnár, G. (2021). Analyzing contextual data in the educational context: educational data mining and logfile analyses. *Journal of Critical Reviews*. 2021; 8(1): 261-273. doi: 10.31838/jcr.08.01.31 .

Alzoubi, O., Fossati, D., Di Eugenio, B., Green, N., & Chen, L. (2013). Predicting students' performance and problem solving behavior from iList log data. In ICCE 2013, *21st International Conference on Computers in Education.*

Bennett, R. E. (2003). *Online assessment and the comparability of score meaning*. Princeton, NJ: Educational Testing Service.

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school-readiness and reasoning skills. *Journal of Educational Psychology*, *106*(2), 639–650.

Dahiya, V. (2018). A survey on educational data mining. *International Journal of Research in Humanities, Arts and Literature, 6*(5), 23–30.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research, 68*(2), 179-201.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, *5*(1).

Dörner, D. & Funke, J. (2017). Complex Problem Solving: What It Is and What It Is Not. *Frontiers in Psychology, 8*:1153. doi: 10.3389/fpsyg.2017.01153

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning, 36*(6), 933–956.

Fischer, F., & Greiff, S. (2021). Assessing complex problem-solving skills in under 20 minutes. *Psychological Test Adaptation and Development*.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning,* 7(1), 69–89

Funke, J. (2021). It requires more than intelligence to solve consequential world problems. *Journal of Intelligence*, 9(3), 38.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, 106(3), 608–626.

Greiff, S. & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. *Zeitschrift für Pädagogik, 56,* 216–227.

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *Journal of Problem Solving, 5*, 71-91. https://doi.org/10.7771/1932-6246.1153

Greiff, S., Krkovic, K., & Hautamäki, J. (2015). The prediction of problem-solving assessed via microworlds. *European Journal of Psychological Assessment, 32*, 298–306.

Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration strategies in computer-simulated complex problem environments: A latent class approach. *Computers & Education*, 126, 248–263.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, *36*(3), 189–213. https://doi.org/10.1177/0146621612439620

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke. J., & Csapó, B. (2013). Complex problem solving in educational contexts – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364–379. doi: 10.1037/a0031856.

Molnár, G., & Csapó, B. (2018). The efficacy and development of students' problem-solving strategies during compulsory schooling: Log-file analyses. *Frontiers in Psychology, 9,* 302.

Molnár, G., & Csapó, B. (2019). How to Make Learning Visible through Technology: The eDia-Online Diagnostic Assessment System. In *Proceedings of the 11th International Conference on Computer Supported Education* (CSEDU 2019) - Volume 2, pages 122–131. ISBN: 978-989-758-367-4.

Naumann, J. (2019). The skilled, the knowledgeable, and the motivated: Investigating the strategic allocation of time on task in a computer-based assessment. *Frontiers in Psychology, 10*, 1429.

OECD. (2014). *Results: creative problem solving - students' skills in tackling real-life problems* (Vol. V). Paris: OECD Publishing.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2021). Measuring collaborative problem solving: research agenda and assessment instrument. *Interactive Learning Environments*, 1-21.).

Reeff, J. P., Zabal, A., & Blech, C. (2006). *The assessment of problem-solving competencies: A draft version of a general framework*. German Institute for Adult Education (DIE).

Schnotz, W., Baadte, C., & Müller, A. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. De Corte, T. De Jong, & J. Elen (Eds.), *Use of representations in reasoning and problem solving* (pp. 21–45).

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42-52.

Stadler, M., Hofer, S. & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior, 111*. https://doi.org/10.1016/j.chb.2020.106442

Wu, H. & Molnár, G. (2021). Logfile analyses of successful and unsuccessful strategy use in complex problem-solving: A cross-national comparison study. *European Journal of Psychology of Education*, 1-24. https://doi.org/10.1007/s10212-020-00516-y

Publications related to the dissertation

Alrababah, S. A. & Molnár, G. (2021). Analysing Contextual Data in Educational Context: Educational Data Mining and Logfile Analyses. *Journal of Critical Reviews*, 8(1), 261-273.

Alrababah, S., Wu, H., & Molnár, G. (2022). Measuring Complex Problem-Solving in Jordan: Feasibility, Construct Validity and Behaviour Pattern Analyses. *Advance*. Preprint. https://doi.org/10.31124/advance.20272437.v1

Alrababah, S. A., Wu, H. & Molnár, G. (2022). Measuring Complex Problem-Solving in Jordan: Feasibility, Construct Validity and Behaviour Pattern Analyses. *SAGE open.* (Submitted).

Alrababah, S. A. & Molnár, G. (2021). The Evolution of Technology-based Assessment: Past, Present, and Future. *International Journal of Learning Technology*. 16(2), 134-157.

Alrababah, S. A. & Molnár, G. (2021, June). Assessing Complex Problem Solving in Jordanian Higher Education Context. Paper presented at inclusive excellence and inclusive universities Conference. Pecs, Hungary.

Alrababah, S. A. & Molnár, G. (2021, November). Jordanian students' test-taking behaviour and problem-solving achievement. Paper presented on the 21st Conference on educational sciences, Szeged, Hungary.

Alrababah, S. Molnár, G. (2019, November). The developmental tendencies of educational assessment: from traditional to third-generation computer-based assessment. Paper presented at XIX. Országos Neveléstudományi Konferencia, Pecs, Hungary

Alrababah, S., Wu, H. & Molnár, G. (2020, December). The Efficacy of Students' Problem-Solving Strategies in Higher Education in Jordan: Log File Analyses. Paper presented at the EARLI SIG 27 Conference. Antwerp, Belgium.

Molnár, G., Alrababah, S. A., & Greiff, S. (2022). How We Explore, Interpret, and Solve Complex Problems: A Cross-National Study of Problem-Solving Processes. *Heliyon*, 8 (1) e08775.

Alrababah, S. Molnár, G. (2022, April). Measuring complex problem-solving in Jordanian higher education: The effect of demographic, cognitive and affective factors on students' achievement. Paper accepted in CEA 2022 conference, Szeged, Hungary.