Measuring Complex Problem-Solving in Jordan: Feasibility, Construct Validity and Behaviour Pattern Analyses

SALEH ALRABABAH, Hao Wu, and Gyöngyvér Molnár

1Affiliation not available

July 25, 2022

Abstract

This study investigates the role of strategic exploration and different problem-solving and test-taking behaviours in CPS success, using logfile data to visualize and quantify students’ problem-solving behaviour on ten CPS problems with different levels of difficulty and characteristics. Additionally, in the present study, we go beyond the limits of most studies that focus on students’ problem-solving behaviour pattern analyses in European cultures and education systems to examine Arabic students’ CPS behaviour. Results show that students in the Arabic school system interpret CPS problems the same way. That is, we confirmed the two-dimensional model of CPS, indicating the processes of knowledge acquisition and knowledge application as separate dimensions during the problem-solving process. Large differences were identified in the test-taking behaviour of students in terms of the efficacy of their exploration strategy. We identified four latent classes based on the students’ exploration strategy behaviour.
Measuring Complex Problem-Solving in Jordan: Feasibility, Construct Validity and Behaviour Pattern Analyses

Saleh Ahmad Alrababah, Hao Wu, and Gyöngyvér Molnár

Abstract

In the 21st century, complex problem-solving (CPS) serves as a key indicator of educational achievement. However, the elements of successful complex problem-solving have not yet been fully explored. This study investigates the role of strategic exploration and different problem-solving and test-taking behaviours in CPS success, using logfile data to visualize and quantify students’ problem-solving behaviour on ten CPS problems with different levels of difficulty and characteristics. Additionally, in the present study, we go beyond the limits of most studies that focus on students’ problem-solving behaviour pattern analyses in European cultures and education systems to examine Arabic students’ CPS behaviour. Results show that students in the Arabic school system interpret CPS problems the same way. That is, we confirmed the two-dimensional model of CPS, indicating the processes of knowledge acquisition and knowledge application as separate dimensions during the problem-solving process. Large differences were identified in the test-taking behaviour of students in terms of the efficacy of their exploration strategy. We identified four latent classes based on the students’ exploration strategy behaviour. The study thus leads to a better understanding of how students solve problems and behave during the problem-solving process in uncertain situations.

Keywords: Complex problem-solving, Logfile analysis, Test-taking behaviour, Higher education, Exploration strategy
1. Introduction

Nowadays, schools should prepare their students for jobs and technologies that do not yet exist and solve problems that have never been faced before (OECD, 2018) to succeed in this new world. Those prospects represent novel needs in higher education and have led to a growing interest in assessment instruments that cover a broader area of competencies than traditional domain-specific skills and disciplinary knowledge (Molnár & Csapó, 2018). These assessment instruments can be used to measure students’ 21st-century skills.

The present study focuses on problem-solving, especially complex problem-solving (CPS), in an Arabic higher education environment. We assess the suitability of education programs in terms of the development of students’ 21st-century skills in Jordan, obtaining more knowledge about the factors and mechanisms that constitute a successful complex problem-solver, while considering the problem-solving behaviour of students socialized in different cultures. This study therefore investigates different factors, such as the role of strategic exploration and test-taking behaviour in CPS success, using logfile data to visualize and quantify students’ problem-solving behaviour in ten CPS problems with different levels of difficulty and characteristics. Our paper is among the first to study the feasibility and validity of an interactive, innovative third-generation computer-based test, such as the globally examined CPS assessments (see e.g. Csapó & Funke, 2017; Dörner & Funke, 2017; Molnár & Csapó, 2018; Molnár, Greiff, Wüstenberg, & Fischer, 2017; OECD, 2014a; Wüstenberg et al., 2014; Wu & Molnár, 2021) in Jordanian higher education. We investigated whether Jordanian students interpret problems the same way as students in other, mostly European, countries (Greiff et al., 2013; Greiff, Wüstenberg, & Avvisati, 2015; OECD, 2014a; Wüstenberg et al., 2014), where most CPS studies have been carried out (Molnár, Alrababah, & Greiff, 2022). Beyond students’ CPS performance, computer-based assessment makes it possible to monitor additional test-taking behavioural actions, such as mouse clicks, time-on-task and problem-solving strategy (Gnaldi, Bacci, Kunze & Greiff, 2020). These sorts of information have the potential to provide policymakers and researchers with valuable insights into students’ CPS skills and offer new ways to assist them in optimizing their cognitive capacity (Wu & Molnár, 2021). Such information is still missing from the evaluation and development of different education systems with increasing cultural diversity.

2. Theoretical background

2.1 Complex problem-solving and its assessment

The classical view defines problem-solving as a step-by-step process, which is passive, reproductive and domain-general, mostly based on trial and error (Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013). In contrast, the Gestalt view considers problem-solving as a productive and active process, where insight, reorganisation and functional fixedness play an important role (Baadte & Müller, 2010). The development of the information-processing approach and Newell and Simon’s problem space theory has opened the door to new directions in research. North American research has typically focused on examining the development of expertise in separate domains, while most of the research in Europe has concentrated on the problem-solving processes of complex, unknown problems with the help of computerized scenarios. Reeff, Zabal, and Blech (2006) defined problem-solving as guided thinking and action
in situations with no routine solution. Eichmann, Goldhammer, Greiff, Brandhuber, and Naumann (2020) distinguished analytical and interactive problem-solving according to the interactive nature of the problem scenario.

In analytical (static) problem-solving environments, both the problem and the related information are static. That is, there are no changes during the problem-solving process, with all the relevant information being presented at the beginning of the problem-solving process (Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013).

Complex problem-solving (CPS) requires a sequence of complex cognitive processes or continuous activities (Funke, 2010). Previous research has recognised two different approaches to measuring CPS (Buchner 1995; Funke, 2014):

1. Computer-simulated microworlds, which have a large number of variables like real-life problems. For example, the well-known microworld scenario “Lohhausen”, which consists of nearly 2000 associated variables (Dörner et al., 1983). This approach results in highly complex problems with high-level similarities to real-world problems. However, (a) their application requires a very long testing time, and (b) they fail to employ common theoretical frameworks to produce comparable problems in a systematic way (Funke, 2001; Funke & Frensch, 2007). In addition, (c) participants’ performance is influenced by many other factors, such as prior knowledge about the problem context, not only their problem-solving skills (Greiff et al., 2015). Finally, (d) the majority of microworld-based problems consist of a few items or many interconnected items, both harming instrument reliability (Greiff et al., 2015).

2. Simplified, artificial, but still complex problems that follow specific construction rules. Most (though not all) of the characteristics of a complex system are present in minimal complex systems (dynamic, complex and intransparent; see Funke, 1991). A minimal complex system has a low number of variables and relations, resulting in reduced testing time compared to the highly complex and challenging microworlds. The MicroDYN approach falls into this category (Greiff & Funke, 2009; Greiff et al., 2012, Schweizer et al., 2013). It applies a number of independent “fake” scenarios to prevent the influence of participants’ previous knowledge (Greiff et al., 2015), it uses only a few variables – that is, problems are easy to scale – and it is widely accepted among problem-solving assessments (see e.g. Csapó & Molnár, 2017; Greiff et al., 2015a; Greiff & Wüstenberg, 2014; Mustafić et al., 2019; OECD, 2014). However, there are limitations to generalization to consider as regards problems of minimal complexity in comparing real-life problems because variables cannot be selectively controlled in a real-life context in most cases (Funke, 2021).

The focus of the present study is on complex problem-solving, especially the MicroDYN approach (Funke, 2014), measured in a computerized and interactive environment. According to the theoretical understanding, complex problem-solving (CPS) in the MicroDYN approach is a two-dimensional construct (Funke, 2001; Greiff et al., 2012; Greiff et al., 2013; Leutner, Wirth, Klieme, & Funke, 2005), consisting of knowledge acquisition and knowledge application. In the first phase of the problem-solving process, the problem-solver needs to acquire knowledge in uncertain situations (knowledge acquisition), while in the second phase, this newly acquired knowledge must be applied in a goal-directed way toward the problem solution (Funke, 2001; Greiff et al., 2018; Novick & Bassok, 2005). In a real-life setting, these two processes are related and take place at the same time. However, in an assessment situation, they are usually separated.
2.2. The role of strategic exploration in problem-solving

Exploring and generating effective information represent the secret to solving a problem successfully. According to Wittmann and Hattrup (2004), “riskier strategies [create] a learning environment with greater opportunities to discover and master the rules and boundaries [of the problem]” (p. 406). Thus, there may be differences in the efficacy of the exploration strategies when gathering information about a problem (Wu & Molnár, 2021). Problem-solvers are supposed to explore the problem environment by acquiring knowledge during strategic exploration (Fisher, Greiff, & Funke, 2012). The development and implementation of strategic exploration are central actions of the problem-solving process (Wüstenberg, Greiff, & Funke, 2012). Problem-solving success in MicroDYN scenarios, which are simplifications and simulations of real-world problems, is also affected by the adoption and application of strategic exploration. In these artificial problem situations, the isolated variation strategy has been the most frequently discussed exploration strategy (it is often called the vary-one-thing-at-a-time strategy; VOTAT; Vollmeyer, Burns, & Holyoak, 1996). Using the VOTAT strategy, the problem-solver directly detects the effects of a single variable at a time by manipulating a given variable in a systematic way, while keeping the other variables unchanged, i.e. in the neutral position (Molnár & Csapó, 2018). According to previous studies, participants who know how to apply VOTAT are more likely to achieve better on problem-solving tasks (Greiff, Molnár, Martin, Zimmermann, & Csapó, 2018), particularly in minimal complex systems (Fischer et al., 2012). According to Lotz, Scherer, Greiff, and Sparfeldt (2017), effective use of VOTAT correlates with higher levels of intelligence, and successful exploration behaviour may lead to better results in problem-solving (Wu & Molnár, 2021).

VOTAT is among the most effective exploration strategies in most problem-solving environments (Lotz et al., 2017; Wu & Molnár, 2021), and it is the most effective in minimal complex systems (such as MicroDYN). Based on Greiff et al. (2018) and Molnár and Csapó (2018), we have discerned and quantified three types of exploration strategies in each of the problem scenarios in the present analyses: (1) No VOTAT (no VOTAT trial was applied); (2) partial VOTAT (VOTAT trials were used for some but not all of the variables in a given problem scenario); (3) full VOTAT (VOTAT trials were applied for all of the variables in a given CPS scenario) (see Greiff et al., 2018; Molnár & Csapó, 2018; Wu & Molnár, 2021).

3. Aims

Nowadays, there is a positive attitude toward using technology in higher education in Jordan (Al-Khayat, 2017), but we do not have any proof of its feasibility and applicability, especially in the field of assessment. Thus, at the initial phase of the study, we had to test the feasibility and applicability of using innovative, interactive, third-generation computer-based tests in Jordan in an educational context. We also explored students’ test-taking and problem-solving behaviour while solving complex problems in a digital environment with both directly collected answer data and logfile analyses. We thus aim:

(Research Aim 1) to test the applicability of an interactive, innovative third-generation test, such as the CPS test, in a country where computer-based assessment has a relatively short history;
(Research Aim 2) to test the structure of the assessed construct (construct validity), that is, to test the underlying dimensionality of CPS measured in the Jordanian educational context, assuming – based on theory and international (mostly European) assessments – a measurement model consisting of two different but highly correlated dimensions or processes of problem-solving (i.e. knowledge application and knowledge acquisition);

(Research Aim 3) to discover and describe the type of strategic exploration used by Jordanian university students while solving CPS problems with different characteristics to understand the mechanism underlying successful CPS; and

(Research Aim 4) to detect the relationships between different types of test-taking and problem-solving behaviour and CPS performance to find new ways to assist students in optimizing their cognitive capacity.

4. Methods

4.1 Participants

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

4.2 Instruments

CPS was measured with a computer-based test developed within the MicroDYN approach (Greiff & Funke, 2017) and adapted into the Arabic style. In MicroDYN, problem environments consist of up to six variables with up to four different type of relations. The problems are embedded in fictitious cover stories, thus eliminating the influence of prior knowledge (for example, “When you get home in the evening, a young cat is lying on your doorstep. It is exhausted and can barely move. You decide to feed the cat. A neighbour gives you two kinds of cat food. Find the relation between the cat food and the cat’s movement/purring”).

The test consisted of six complex problems with different characteristics and different levels of complexity. On each MicroDYN problem, participants were first expected to explore the structure of the problem scenario by freely operating the system during the knowledge acquisition phase, that is, by manipulating one or more input variables (displayed on the right side according to the Arabic style) for no more than three minutes (see Fig. 1), and then analyse their effects on the output variables (displayed on the left side according to the Arabic style). In parallel, within the 180 seconds of the knowledge acquisition phase, they were expected to visualize the detected relations by drawing lines between the variables on a concept map presented at the bottom of the screen (see Fig. 2). The history of the settings was shown on a graph linked to each input and output variable. In practice, each problem scenario has four buttons beyond the adjustment sliders and buttons for the input variables: Help, Apply, Reset and Next. By clicking on the Reset button, the participant has the option of deleting all the histories presented on the graphs and setting all the values back to their original values. Each input variable has five stages: +2 (++), +1 (+), 0, −1 (−) and −2 (−−), which can be set using the sliders or buttons (+ or −) next to the input variables. Their effects on the output variables can be tested by clicking on the Apply button. The changes in the output variables are presented in both numerical
and graphic formats in the problem scenario. The Next button makes it possible to navigate between the MicroDYN scenarios and its different phases.

Fig. 1. Sample item from the Arabic-language version of the CPS test – Knowledge acquisition phase. In the example, the task is to find out about the effects of sport and reading on endurance and strength. The controllers of the input variables range from “- -” (value=-2) to “++” (value=+2). In the English version (to the right), they are presented on the left side of the problem environment and on the right side in the Arabic version (to the left). The model is shown at the bottom of the figure.

Fig. 2. Example of problem representation: Drawing relations on a concept map provided onscreen. The English version is provided to the right.

Second, in the knowledge application phase, students are expected to use the system in a goal-directed way to reach particular target values (e.g. a given level of movement/purring) of the output variables. To avoid item dependence in this phase, the right concept map is presented at the bottom of the screen. In this part of the problem-solving process, students have no more than 90 seconds and four trials (clicking four times on the Apply button) to solve the problem, that is, to reach the target values of the output variables. Fig. 3 provides a screenshot of the knowledge application phase for a problem with four variables (two input and two output variables) with two direct effects.
The items were adapted from the European to the Arabic writing style by changing the direction of the items from left to right to make them suitable for the right-to-left reading and writing convention of the Arabic language (see Figs. 1 and 3) and by translating the instructions into Arabic. The complexity of the problem was scaled by the number of variables (input-output; 2-2, 3-2, 3-3) and the number (2-4) and type (direct or indirect) of relations. According to Beckmann, Birney, and Goode (2017), raising the number of both variables and relations will boost the difficulty of the CPS problems.

4.3 Procedure

Test administration. The eDia online assessment platform (Molnár & Csapó, 2019) was used for the test administration. The data collection lasted 45 minutes at each university’s computer labs. As an achievement indicator, we applied the traditional scoring for both CPS phases (see e.g. Csapó & Molnár, 2017; Fischer et al., 2012; Molnár & Csapó, 2018):

Scoring the answers and labelling the logfiles. If the visualized relations matched the theoretical structure of the problem, students obtained a score of 1. Otherwise, the response was assigned 0 points (for the first phase). Further, if the problem-solver managed to achieve the target values of the output variables within the given time (90 min.) and trial frames (clicking on the Apply button four times), students earned another 1 point, or 0 points otherwise. Applying the traditional scoring, we generated databases for the analyses for Research Aims 1 and 2. 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 labelled in the databases. Students’ problem-solving behaviour was defined in each problem.
situation separately by evaluating all of the trials executed within the same problem. If the problem-solving behaviour followed 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 used only for some, but not for all of the input variables – which was assigned a score of 1 point; and (c) full VOTAT, when the VOTAT strategy was used for all the input variables – which garnered a score of 2 points. These scores provided the foundation for fulfilling Research Aims 3 and 4.

Data analyses. The descriptive analyses were executed by SPSS (Research Aim 1). Confirmatory factor analyses was used to test the underlying measurement model of complex problem-solving, assuming two different problem-solving processes, knowledge acquisition and application. These analyses were executed by MPlus (Research Aim 2). We have accepted the cut-off values suggested by Hu and Bentler (1999), who indicated that a CFI (Comparative Fit Index) and a TLI (Tucker–Lewis Index) value above .95 and a RMSEA (Root Mean Square Error of Approximation) below .06 indicate a good model fit. We used the preferred estimator for categorical variables, Weighted Least Squares Mean and Variance adjusted (WLSMV; Muthén & Muthén, 2010). Latent class analysis (LCA) was used for Research Aim 3 and was also executed by MPlus. LCA is a pattern-finding algorithm, which searches for latent classes which share similarly observed variables (Collins & Lanza, 2010). In this study, LCA was used to establish latent classes regarding students’ problem-solving behaviour. The quality of the LCA was evaluated with the following fit indices: the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and adjusted Bayesian Information Criterion (aBIC). As regards these fit indices, lower values indicate a better model fit. Entropy was utilized to test the accuracy of the classification. The Lo–Mendell–Rubin Adjusted Likelihood Ratio was used to compare the LCA models with different numbers of latent classes (Lo et al., 2001).

5. Results

5.1. Results for Research Aim 1 focusing on testing the applicability of an interactive, innovative third-generation test, such as the CPS test, in a country where computer-based assessment has a short history

Using the traditional scoring for the CPS problems, the internal consistency of the test was high (Cronbach’s alpha=.83). The phase-level reliabilities also proved to be good and acceptable (KAC (knowledge acquisition phase): .83; KAP (knowledge application phase): .65). The test proved to be difficult for the students (M=16.8%; SD=16.7% points), whose achievement was significantly higher in the knowledge acquisition phase (M=25.3%; SD=25.7% points) than in the knowledge application phase (M=8.1%; SD=13.0% points; t=10.2, p<.001). To sum up, using interactive, innovative, third-generation computer-based assessments is feasible and reliable in the Jordanian higher education context.

5.2 Results for Research Aim 2 focusing on the underlying dimensionality of CPS measured in the Jordanian educational context, assuming – based on theory and international assessments – a measurement model with two different problem-solving processes (i.e. knowledge application and knowledge acquisition)

The bivariate correlations between the two CPS processes, knowledge acquisition and knowledge application, proved to be medium (r=.45; see Table 1), indicating the measurement of different aspects of CPS.
Table 1. Test and phase level correlations

<table>
<thead>
<tr>
<th></th>
<th>KAC</th>
<th>KAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAC</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>KAP</td>
<td>.453*</td>
<td>1.00</td>
</tr>
<tr>
<td>CPS</td>
<td>.925*</td>
<td>.758*</td>
</tr>
</tbody>
</table>

Note: KAC: knowledge acquisition; KAP: knowledge application; CPS: complex problem-solving; **p<.01 level significant

Confirmatory factor analyses indicated a good fit (see Table 2). A special $\chi^2$-difference test in Mplus (Muthén & Muthén, 2010) was carried out to compare the one- and two-dimensional models. This test revealed that the two-dimensional model fit the data significantly better (Chi-Square Test for Difference Testing=55.317, df=1, p<.001). Thus, we confirmed the theory and the earlier empirical results based on European and Asian data collections as regards CPS (Wu & Molnár, 2021). CPS is a two-dimensional construct, where the KAC and KAP processes can be distinguished empirically.

Table 2. Goodness of fit indices for testing dimensionality of CPS in Jordan

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-dimensional</td>
<td>234.938</td>
<td>169</td>
<td>.001</td>
<td>.965</td>
<td>.961</td>
<td>.045</td>
</tr>
<tr>
<td>1-dimensional</td>
<td>289.945</td>
<td>170</td>
<td>.001</td>
<td>.936</td>
<td>.929</td>
<td>.061</td>
</tr>
</tbody>
</table>

Note. df=degrees of freedom; CFI=Comparative Fit Index; TLI=Tucker–Lewis Index; RMSEA=Root Mean Square Error of Approximation; WLSMV estimator was used in the analyses.

5.3. Results for Research Aim 3 to discover and describe the exploration behaviour of the Jordanian university students while solving computer-based CPS problems with different characteristics

Contrary to our expectations, based on the results for Research Aim 1, the percentage of theoretically effective strategy use was 56.5% for the more complex problems and 64.2% for the less complex ones (see Table 3).

Table 3. Percentage of theoretically effective and non-effective strategy use

<table>
<thead>
<tr>
<th>Complexity of problem</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of input and output variables</td>
<td>Number of relations</td>
</tr>
<tr>
<td>2-2</td>
<td>2</td>
</tr>
<tr>
<td>3-2</td>
<td>3</td>
</tr>
<tr>
<td>3-3</td>
<td>4</td>
</tr>
</tbody>
</table>

A large percentage of the Jordanian students employed theoretically effective exploration strategies, including the VOTAT strategy, where the problem-solver manipulates only one input variable systematically while at the same time keeping the other variables unchanged to be able to test the direct effect of the input variables under investigation on the output variables during the problem-solving process. These manipulations allow direct monitoring of changes in output variables to demonstrate the impact of the variable just modified (Molnár & Csapó, 2018). Table 4 summarizes the percentage of no VOTAT, partial VOTAT and full VOTAT strategy users. Independently of problem complexity,
a majority of the students applied the most effective exploration strategy during the problem-solving process, but, according to the results for Research Aim 1, they were unable to interpret its meaning. That is, at the very end, most of them failed to solve the problems properly.

**Table 4.** Percentage of no VOTAT, partial VOTAT and full VOTAT strategy use

<table>
<thead>
<tr>
<th>Complexity of problem</th>
<th>Number of input and output variables</th>
<th>Number of relations</th>
<th>No VOTAT (%)</th>
<th>Partial VOTAT (%)</th>
<th>Full VOTAT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2-2</td>
<td>3</td>
<td>52.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3-2</td>
<td>4</td>
<td>56.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3-3</td>
<td>6.9</td>
<td>53.1</td>
</tr>
</tbody>
</table>

**5.4. Results for Research Aim 4 to detect the relationships between different types of problem-solving behaviour and problem-solving performance**

Only half (52.1%) of the students who applied a theoretically correct strategy made a correct decision as well, solving the easiest problems correctly. This ratio increased to 59.8% with the second sort of complexity before dropping slightly on the most complex problems. Note that the complexity of a problem was defined by the number of variables and the number of relations (Table 5).

**Table 5.** The ratio of high and low achievers among the theoretically effective strategy users during problem-solving

<table>
<thead>
<tr>
<th>Complexity of problem</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of input and output variables</td>
<td>Number of relations</td>
</tr>
<tr>
<td>2-2</td>
<td>2</td>
</tr>
<tr>
<td>3-2</td>
<td>3</td>
</tr>
<tr>
<td>3-3</td>
<td>4</td>
</tr>
</tbody>
</table>

Fig. 4 displays the ratio of high and low achievers among theoretically effective strategy users at the task level. The ratio for the effective strategy users to correctly solve an item was higher than 50% on most of the items (except on item 2). Compared to the relatively low performance for the overall sample (see Section 5.1), the theoretically effective strategy users showed a remarkably better performance.

Problem-solving performance among the theoretically effective strategy users suggests the guessing factor, which indicates a correct solution despite theoretically non-effective strategy usage. This also includes participants who recall a theoretically effective strategy but apply it wrongly and then solve the problem (see Table 6). The guessing factor varied from 15.3% to 7.5%, from the least to the most complex tasks.

The guessing factor (indicating those who solve a problem without an effective strategy) showed the highest effectiveness on item 1. Item 1 is of the 2-2 type, which is the easiest. The effectiveness dropped for the rest of the items. Low achievement for the non-theoretically effective strategy users was very noticeable for all the CPS items.
Fig. 4. Problem-solving performance among the theoretically effective strategy users

Table 6. Problem-solving effectiveness among the theoretically non-effective strategy users

<table>
<thead>
<tr>
<th>Complexity of problem</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of input and output variables</td>
<td>Number of relations</td>
</tr>
<tr>
<td>2-2</td>
<td>2</td>
</tr>
<tr>
<td>3-2</td>
<td>3</td>
</tr>
<tr>
<td>3-3</td>
<td>4</td>
</tr>
</tbody>
</table>

Fig. 5. Problem-solving performance among the theoretically non-effective strategy users

After analysing the performance of theoretically right and theoretically wrong strategy users, we went further to obtain a statistical model of students’ problem-solving ability. First, using the tools of latent class analysis and log data to ascertain the use of VOTAT strategies based on students’ exploration behaviour, we distinguished three
qualitatively different VOTAT strategy users. The Akaike, Bayesian and adjusted Bayesian Information Criterion indices decreased with a growing number of latent classes up to the 4-class solutions. The entropy index reached its maximum value for the 2-class model. However, it was also high for the 3- and 4-class solutions. The Lo–Mendell–Rubin adjusted likelihood ratio test indicated the best model fit for the 3-class model, and it proved to be no longer significant for the 4-class model. Thus, we used the 3-class model – where 93% of the Jordanian students were accurately categorized – to distinguish three qualitatively different class profiles in the further analyses: 50.5% of these students were among the proficient strategy users, who consistently employed VOTAT strategies almost from the very first problem; 18.1% proved to be intermediate explorers, who used VOTAT strategies with lower but still intermediate frequency; and 31.4% were low-level strategy users, who barely made use of VOTAT strategies throughout the assessment process.

**Table 7.** Fit indices for latent class analyses monitoring students’ problem-solving behaviour in uncertain situations

<table>
<thead>
<tr>
<th>Number of latent classes</th>
<th>AIC</th>
<th>BIC</th>
<th>aBIC</th>
<th>Entropy</th>
<th>L–M–R test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1504</td>
<td>1585</td>
<td>1506</td>
<td>0.967</td>
<td>613</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3</td>
<td>1446</td>
<td>1570</td>
<td>1449</td>
<td>0.931</td>
<td>83</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>4</td>
<td>1440</td>
<td>1607</td>
<td>1445</td>
<td>0.944</td>
<td>31</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>

Table 8 indicates the problem-solving performance of all three classes of participants (low-level strategy users, intermediate explorers and expert explorers). The results indicate that all three classes of participants performed better on the easier items (the 2-1 and 2-2 types) than on the more complex problems (the 3-3 type). Furthermore, the results confirmed that VOTAT is the most effective strategy. Problem-solvers that used it had a higher chance to solve a problem correctly, with the exception that the intermediate explorers performed slightly worse than the low-level strategy users on the 3-3 problems.

**Table 8.** Problem-solving performance for low-level strategy users, intermediate explorers and expert explorers

<table>
<thead>
<tr>
<th>Latent class profiles</th>
<th>Frequency (%)</th>
<th>2-2 problems</th>
<th>2-3 problems</th>
<th>3-3 problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>High achievement</td>
<td>Low achievement</td>
<td>High achievement</td>
</tr>
<tr>
<td>Low-level strategy users</td>
<td>32.0</td>
<td>68.0</td>
<td>34.4</td>
<td>65.6</td>
</tr>
<tr>
<td>Intermediate explorers</td>
<td>33.8</td>
<td>66.2</td>
<td>35.3</td>
<td>64.7</td>
</tr>
<tr>
<td>Expert explorers</td>
<td>45.4</td>
<td>54.6</td>
<td>46.5</td>
<td>53.5</td>
</tr>
</tbody>
</table>

6. Discussion

_Research Aim 1: To test the applicability of an interactive, innovative third-generation test, such as the CPS test, in a country where computer-based assessment has a short history._
In this study, we used logfile analysis to examine Jordanian undergraduate students’ problem-solving behaviour. First, we monitored the feasibility and applicability of computer-based assessment in the Jordanian educational context. The internal consistency of the CPS tests was high, but the mean achievement was relatively low, indicating that it is difficult for the students to solve interactive problems. Based on all the descriptive results, we can conclude that computer-based assessment and innovative online tests are feasible and valid in Jordan at the level and in the context of higher education.

Research Aim 2: To test the structure of the assessed construct (construct validity), that is, to test the underlying dimensionality of CPS measured in the Jordanian educational context, assuming – based on theory and international (mostly European) assessments – a measurement model with two different problem-solving processes (i.e. knowledge application and knowledge acquisition).

The analyses of the structural stability of the measured construct confirmed earlier research results obtained in Europe (e.g. Funke, 2001; Wüstenberg, Greiff, & Funke, 2012) and Asia that CPS is a two-dimensional construct. The processes of the KAC and KAP phases can also be empirically distinguished in the Jordanian context. The bivariate correlation (r=.45) between KAC and KAP was consistent with earlier research results, which varied between r=0.14 and r=0.94 (Nicolay, Krieger, Stadler, Gobert, & Greiff, 2021). The reason for this wide range of correlation coefficients is the use of different problem-solving approaches and CPS assessments to measure KAC and KAP. Since CPS skills are a key competence for educational success, research results on CPS have important implications for filling the gap between students’ ability to acquire and then apply that knowledge in uncertain situations, which has become highly important in the 21st century. CPS serves as a relevant showcase for addressing a crucial existing gap in modern-day education: the gap between students’ ability to acquire knowledge and then apply this knowledge in uncertain situations, which is increasingly significant in the 21st century.

Research Aim 3: To discover and describe the type of strategic exploration used by Jordanian university students while solving CPS problems with different characteristics to understand the mechanism underlying successful CPS.

Logfile-based analyses have expanded the scope of previous studies on CPS, especially in the Arabic environment, and enabled us to identify key components of students’ problem-solving skills: the way they explore and understand relatively simplified problems and the relationships within the problem. A large number of students showed systematic strategies but failed to solve the problem; that is, the use of a theoretically effective strategy does not always lead to high problem-solving achievement, a finding which confirms research results by de Jong and van Joolingen (1998), who claim that learners often have trouble understanding data. In contrast, we have detected another relatively large number of students who achieved high performance without collecting all the information necessary to be able to solve the problem correctly; that is, they applied a theoretically non-effective exploration strategy. Beyond guessing, it is more difficult to find a clear explanation for this discrepancy in students’ problem-solving behaviour. The result is consistent with previous research (e.g. Greiff et al., 2015; Molnár & Csapó, 2018; Vollmeyer et al., 1996) that indicates that high performance is not always in line with the right kind of problem exploration and interpretation. To sum up, the use of a theoretically effective strategy does not always lead to high performance, and, in contrast, high
performance does not always indicate the right kind of exploration and interpretation, i.e. the application of the right kind of problem-solving strategy.

Research Aim 4: To detect the relationships between different types of test-taking and problem-solving behaviour and CPS performance to find new ways to assist students in optimizing their cognitive capacity.

The analysis explored Jordanian students’ problem-solving behaviours in greater depth, focusing on the type of problem exploration and helping us to understand the reasons behind discrepancies between the high ratio of theoretically right exploration behaviour, i.e. collecting information, and low problem-solving achievement. One possible explanation is that students did not provide the proper meaning for the information obtained during the first phase of the problem-solving process. Molnár and Csapó (2018) have shown that there is an inverse relation between problem complexity and the probability of strong problem-solving performance without the use of an effective problem-solving strategy. It is clear that the achievement for all participants on the easiest item (2-2) was not the best. Students’ performance was better on problems of medium complexity (2-3) because they had sufficient experience after solving the first type of problem. On more complex problems (3-3), students’ performance declined, despite having sufficient experience in solving problems. As regards the increasing numbers of input variables, output variables and relations between them, the participants experienced greater difficulty (Beckmann et al., 2017). More analyses are required to detect the reasons for the large differences between the expertise level in the exploration and the lower achievement in the decisions made in problem-solving.

7. Limitations

The study is considered as a small-scale study with 195 participants from two Jordanian universities. Thus, it does not represent the entire university student population in Jordan. Hence, the results from this study are not generalizable. A bigger sample size from more universities and faculties are required to obtain a wider view of Jordanian students’ problem-solving behaviour.

In addition, some participants suffered from the weakness of the internet during the test at peak intensity; all the students used the university system at the same time. This caused some difficulty in retaining access, as some sessions required a high-speed connection. Another limitation stems from the translation and adaptation of the items. Originally, the languages of the items were German and Hungarian. Then, both the Hungarian and German versions were translated into English. After validating the Hungarian, German and English versions, the test was translated into Arabic by specialist translators for distribution to the Jordanian students. Beyond translating the problem texts and instructions, we changed the direction of the test to suit the Arabic format, from right to left (the earlier versions of the test were produced in left-to-right format). We also changed tables, boxes, pictures and all the connecting elements.

The MicroDYN approach was used in the study to assess students’ problem-solving abilities using an instrument which is valid and reliable for measurement purposes, but uses artificial problems, where the number of variables and relations are limited. Hence, the problem-solving behaviour observed in MicroDYN scenarios cannot be generalized to all types of problems we face in everyday life.
8. Conclusion and implications

The study points to the possibility and feasibility of problem-solving measurements in the Jordanian context. It highlights the importance of explicit development of problem-solving skills and problem-solving strategies as a means of applying knowledge in new contexts in higher education. The findings highlight the importance of developing instructional methods to improve students’ CPS skills by enhancing their individual learning strategies. The results also suggest the need for further investigation to explore a larger representation of the relationships between students’ cognitive skills and their behaviour in problem-solving situations. To sum up, the study has shed light on Jordanian students’ problem-solving development from the perspective of their behaviour, thus providing a solid basis for further study in the Jordanian context.

9. Funding

This research was supported by the Hungarian National Research, Development and Innovation Fund (grant under the OTKA K135727 funding scheme) and the Hungarian Academy of Sciences (Research Programme for Public Education Development of the Hungarian Academy of Sciences, grant KOZOKT2021-16).

10. References


https://doi.org/10.1080/10705519909540118


OECD (2018). The future of education and skills. Education 2030. OECD.


Vo, D. Csapó, B. (2020). Development of inductive reasoning in students across school grade levels. Thinking Skills and Creativity, 37, 100699.


