European Journal of Psychology of Education (2021) 36:1009–1032
https://doi.org/10.1007/s10212-020-00516-y

Logfile analyses of successful and unsuccessful strategy use in complex problem-solving: a cross-national comparison study

Hao Wu[1] · Gyöngyvér Molnár[2]

Received: 24 June 2020 / Revised: 11 October 2020 / Accepted: 12 November 2020 /
Published online: 2 January 2021

Abstract
The purpose of this study is to examine cross-national differences in students' exploration strategies in a computer-simulated CPS (complex problem-solving) environment and to identify similarities and qualitative differences in the way Hungarian and Chinese students explore a CPS environment. In a sample of 187 Chinese and 835 Hungarian students (aged 12), we administered problem-solving items via the eDia platform within the MicroDYN approach. After structuring and coding the logfile data, latent class analyses were used to identify students whose problem-solving strategies showed similar patterns. Results indicated that Chinese students employed the most effective and successful exploration strategy, the VOTAT (vary-one-thing-at-a-time) strategy, more frequently and effectively than Hungarian students and that they showed a significantly higher learning effect during testing than their Hungarian peers. These results highlight the possibilities and importance of explicit enhancement of exploration strategies as a tool for learning in a new technological context.

Keywords Cross-cultural research · Computer-based assessment · Complex problem-solving · Process data

Introduction

Complex problem-solving

Nowadays, the ability to solve problems properly in a timely manner is gradually becoming a key factor in peoples' career and life (OECD 2014). Problem-solving is thus considered to be

✉ Hao Wu
haowu@edu.u–szeged.hu

[1] Research Group on Learning and Instruction, University of Szeged, H-6722 Petőfi Sándor Sgt. 30 –34, Szeged, Hungary

[2] Department of Learning and Instruction, MTA–SZTE Research Group on the Development of Competencies, University of Szeged, Szeged, Hungary

one of the most important twenty-first-century skills because the problems we face are different most of the time. Certain problems are well-defined with clear goals, but sometimes we need to face ill-defined problems which have no clear problem definition (Dörner and Funke 2017) and without "a clear goal, specific investigation boundaries and solving methods" (Lien et al. 2020, p. 134977). Sometimes, the problems are similar or even the same as the problems we have solved before, but it is also entirely possible for us to experience certain problems that we have never encountered before (Frensch and Funke 1995). Some problems are static, while others are dynamically changing (Greiff et al. 2013b; Molnár et al. 2017). Some problems are domain-specific and deeply relevant with separate, natural knowledge domains, but most real-life problems are domain-general and cross-curricular (Frensch and Funke 1995; Greiff et al. 2014b). Both the problem-solving processes and the mental steps can differ widely in solving different types of problems. This study focuses on one specific kind of problem-solving, which is complex problem-solving (CPS) in the MicroDYN approach.

Problem-solving occurs when we want to "overcome barriers between a given state and a desired goal state by means of behavioural and/or cognitive, multistep activities" (Frensch and Funke 1995, p. 18). The state transitions happen in a problem-solver's "problem space"; that is an internal representation of the problem involves the initial state, possible states and the desired goal states of the problem as well as the applicable operators (Newell and Simon 1972). In past decades, the emphasis in problem-solving research gradually shifted from simple, static, well-defined academic problems to more complex, dynamic, ill-defined, real-world problems (Fischer et al. 2012; Wenke et al. 2005). Psychologists further explored and developed the cognitive construct of complex problem-solving on the basis of Newell and Simon's problem space theory. A complex problem refers to a problem situation which is intransparent, changes dynamically and contains numerous elements with a high level of interconnectedness (Frensch and Funke 1995; Greiff et al. 2013b). The problem-solver has to systematically interact with the complex system to generate and integrate information about the problem (Fischer et al. 2012; Funke 2001; Greiff et al. 2013b). The process of complex problem-solving is split into two phases: knowledge acquisition and goal-oriented knowledge application (see e.g. Greiff et al. 2013a; Leutner et al. 2005). In the knowledge acquisition phase, the problem-solver acquires knowledge of the problem, thus establishing representation of the problem space (Funke 2001; Mayer and Wittrock 2006), while he/she applies the acquired knowledge to reach the goal in the knowledge application phase (Funke 2001; Novick and Bassok 2005). Logfile analysis makes it possible to distinguish the exploration and the representation parts of the knowledge acquisition phase (see Molnár and Csapó 2018).

In the CPS environment, it is not an easy task for a problem-solver to construct a parsimonious but effective representation of the problem space in the knowledge acquisition phase. Problem-solvers are assumed to explore the system by applying a strategy (Fischer et al. 2012). Therefore, selection and employment of systematic strategies are important in the CPS process (Molnár and Csapó 2018).

Exploration strategies in a complex problem-solving environment

Problem-solving strategies are mainly for "(a) generating relevant information and (b) making good forecasts and decisions in complex environments" (Fischer et al. 2012, p. 29). In a CPS environment, the information provided for problem-solving is usually insufficient, so the problem-solver has to actively acquire the information by systematically interacting with the system (see Funke 2001). Thus, in the case of CPS, employing efficient exploration strategies to effectively gather information has been considered as one of the key actions in a successful problem-solving process (Wüstenberg et al. 2012; Wüstenberg et al. 2014a).

The isolated variation exploration strategy has frequently been discussed in a number of psychology domains, including CPS (Greiff et al. 2018; Molnár and Csapó 2018; Chen and Klahr 1999; Chen et al. 2019; Kuhn et al. 1995; Wüstenberg et al. 2014a; Schoppek and Fischer 2017). This strategy can also be referred to as the “Vary-One-Thing-At-A-Time” or VOTAT strategy (Vollmeyer et al. 1996; Vollmeyer and Rheinberg 1999). To be more specific, the VOTAT strategy is defined as the method for “systematically [varying] only one input variable, while the others remain unchanged. This way, the effect of the variable that has just been changed can be observed directly by monitoring the changes in the output variables” (Molnár and Csapó 2018, p. 2). Efficiently understanding and controlling the use of VOTAT strategy has been considered as the basis for developing more sophisticated and complex strategies for coordinating multiple variables and the foundation for the phases of scientific thinking (i.e. inquiry, analysis, inference and argument; Kuhn 2010; Kuhn et al. 1995; Mustafić et al. 2019).

The effectiveness of VOTAT in a CPS environment has been frequently discussed. Previous studies have indicated that students who are able to apply VOTAT are more likely to achieve higher performance in a CPS assessment (e.g. Molnár and Csapó 2018; Greiff et al. 2015c). Vollmeyer et al. (1996) pointed out that VOTAT is more effective for acquiring knowledge than other possible strategies (CA: “changing all variables in a haphazard way”; HT: “the heterogeneous collection of all other strategies”, Vollmeyer et al. 1996) in the CPS environment. Later, empirical studies (see Molnár and Csapó 2018; Wüstenberg et al. 2012) confirmed this finding; that is students who are able to apply VOTAT are more likely to achieve higher performance in CPS environments. In the empirical studies, some students’ exploration behaviour, i.e. their use of VOTAT, was consistent throughout the CPS assessment process. However, some students showed lower proficiency in their use of VOTAT when problem complexity increased (see e.g. Molnár and Csapó 2018). In contrast, some students showed a rapid learning effect and a remarkable increasing trend in their proficiency in the use of VOTAT during the testing process (see e.g. Greiff et al. 2018; Molnár and Csapó 2018; Mustafić et al. 2019). This phenomenon is also supported by Kuhn’s study (2012, in the field of scientific reasoning) from a theoretical perspective. Therefore, when analysing and discussing students’ exploration strategy in a CPS environment, it is important to observe students’ behavioural pattern throughout the test process.

Assessing complex problem-solving: computer-based assessment and logfile analysis

ICT (Information and Communications Technologies) refers to a “diverse set of technological tools and resources used to transmit, store, create, share or exchange information” (UNESCO Institute for Statistics 2009, p. 120). It provides numerous opportunities in a range of educational practice, including educational assessment (Csapó et al. 2012). ICT can support computer-based testing to provide a unique assessment environment, where both dynamic and interactive situations are available (Greiff et al. 2014a). It paves the way for measuring problem-solving from a different perspective and applying dynamism, that is interactivity, between the test stimuli and the problem-solver. It allows students to interact with complex simulations and change the items dynamically (Molnár et al. 2017). Indeed, “the inclusion of interactive problems, in which students need to explore the (simulated) environment and gather feedback on the effect of their interventions in order to obtain all the information needed to solve a problem, was only possible by asking students to use a computer to complete the assessment" (OECD 2014, p. 30). In conclusion, computer-based assessment opens doors to studying previously researched knowledge and skill domains in an innovative and new environment or to explore features of twenty-first-century skills, including CPS.

There are two approaches used in CPS measurements (Buchner 1995): (1) computer-simulated microworlds similar to the real world and composed of a large number of variables. For instance, the famous microworld scenario "Lohhausen" contains more than 2000 interconnected variables (Dörner et al. 1983). Due to its high complexity, this approach requires a long testing time. Moreover, the validity of this measurement approach has also been questioned (Greiff et al. 2015a, 2015b). Researchers have argued that microworld-based scenarios focused on a real-world resemblance but failed to use common theoretical frameworks to systematically generate comparable problem-solving tasks (Funke 2001; Funke and Frensch 2007). In this case, not only participants' CPS ability but also their prior knowledge about a given problem situation were reported to have an influence on their performance (Greiff et al. 2015b). Furthermore, most microworld-based problem-solving assessment was based on only one microworld scenario, which consisted of numerous interrelated items. That is, participants' actions in completing each item are not independent. Participants' previous actions will therefore inevitably affect their subsequent actions, thus violating classical testing theory and harming instrument reliability (Greiff et al. 2015b). (2) Simplistic, artificial, but still complex problems following certain construction rules. Minimal complex systems comprise most (or even all) of the features of a complex system (complexity, dynamics, polytely and intransparency; see Funke 1991). Meanwhile, a minimal complex system has low values for the parameters. Thus, it reduces the testing time to a minimum, especially if we compare it to the extremely difficult microworlds. Therefore, the minimal complex system approach (MicroDYN approach; Greiff and Funke 2009; Greiff et al. 2012, Schweizer et al. 2013) is widely accepted and employed to design complex problem-solving assessment (see e.g. Csapó and Molnár 2017; Greiff et al. 2015a; Greiff and Wüstenberg 2014; Mustafić et al. 2019; OECD 2014). Furthermore, the MicroDYN approach also overcomes the major shortcomings of traditional microworlds in that it was designed to avoid influence from participants' prior knowledge and consists of numerous independent scenarios (Greiff et al. 2015b).

In the context of computer-based assessment, students' operations and interactions with the problem and test environment can be recorded, resulting in what we call logfiles (Zoanetti and Griffin 2017) and providing new opportunities to discover and conduct new forms of data analysis in educational assessment. In a CPS environment, logfile data have been used to explore students' problem-solving behaviours (Tóth et al. 2017), exploration strategies (Molnár and Csapó 2018; Chen et al. 2019; Greiff et al. 2015c; Ren et al. 2019), problem-solving proficiency and test-taking motivation (Zoanetti and Griffin 2017), and thus help us to gain a much deeper understanding of how they interact with problems and how they behave during the problem-solving and test-taking process. Using logfile analysis, we focused on their exploration behaviour in the problem-solving process. Students' actions were logged and coded according to the input behaviour model with a labelling procedure. Then, we distinguished whether their exploration strategies fall within the VOTAT scope.

Significance of research

Students with different cultural backgrounds can behave and perform differently in a CPS environment (see e.g. Greiff et al. 2015c; OECD 2014; Wüstenberg et al. 2014b). According to the PISA 2012 problem-solving results, students from East Asian countries showed generally better performance than their mates from European countries (OECD 2014). The results indicated the importance of conducting a problem-solving study in both East Asia and Europe to identify possible reasons for differences and to establish the theoretical and empirical foundations for further training programmes. China (mainland) and Hungary were chosen based on their PISA results (high achiever vs. below average achiever).

Molnár and Csapó's (2018) study has demonstrated that students' exploration strategies influence their problem-solving performance. Therefore, the present study expects that the quality of students' exploration strategies explains their performance and offers a closer look into the reasons for differences in achievement at a certain level.

Very few studies have been published on students' exploration strategies in a problem-solving environment in China. Thus, several questions have currently remained unanswered. According to Molnár et al. (2013), students' reasoning skills (including CPS) start to develop rapidly at around the age of twelve and tend to be stable at around age 15. PISA 2012 has already shown that 15-year-old Chinese students have a higher ability level than their Hungarian contemporaries (OECD 2014). However, the issue of whether Chinese students have better CPS skills than their Hungarian peers in the first phase of the CPS developmental curve (that is 12-year-olds) and whether there is a gender difference in these two countries' students is still unexplored. Moreover, considering that there is a significant cultural difference between China and Hungary, we expect that students will employ different exploration strategies during the problem-solving process which fit their cognitive styles better. Whether the different strategy uses result in different achievement would also be an essential question to be answered. Thus, this study aims to fill these gaps using logfile analysis to examine the similarities and qualitative differences in the way Hungarian and Chinese students explore a problem environment and identify how their strategy use influences their performance in a CPS environment.

Aims and research questions

The purpose of the study is to gain a further understanding of the similarities and differences in Chinese and Hungarian students' behaviour in a complex problem-solving environment. Considering that CPS has rarely been compared between Chinese and Hungarian students, the study aims to firstly analyse if measurement invariance holds, that is if the results of Hungarian and Chinese students can be represented on the same scale. As PISA has indicated a remarkable gap between Chinese and Hungarian 15-year-old students' CPS performance, this study aims to measure if such performance differences exist before the age of 15, in the first phase of the CPS developmental curve (among 12-year-olds; Molnár et al. 2013). Furthermore, the study aims to analyse Hungarian and Chinese students' problem-solving strategies by analysing the similarities and differences between Chinese and Hungarian students' (1) use of exploration strategies (i.e. the proportion of their use of VOTAT); (2) effectiveness of transforming the use of VOTAT to successful problem-solving; and (3) behavioural pattern in their use of VOTAT. Thus, in general, the study seeks to answer five research questions:

(RQ1) Can CPS be measurement invariant across gender and nationality in the contexts of both China and Hungary?

(RQ2) Can developmental differences in CPS be detected between 12-year-old Hungarian and Chinese students? Can a gender difference be detected in the Chinese and Hungarian samples?

(RQ3) Do 12-year-old Chinese and Hungarian students employ different exploration strategies during the problem-solving process? How frequently do they apply the most effective exploration strategy, the VOTAT strategy? Is there a difference between these two groups of students?

(RQ4) Can any differences be detected between Chinese and Hungarian students in the effectiveness of their use of the VOTAT strategy?

(RQ5) What are the similarities and differences in Chinese and Hungarian students' behavioural pattern in their use of VOTAT? What exploration strategy profiles can be detected among Chinese and Hungarian students?

Methods

Participants and procedure

The sample was drawn from sixth-grade, 12-year-old students in Hungarian and Chinese primary schools. One hundred eighty-seven Chinese students (85 boys and 102 girls; mean age = 11.93, SD = 1.06) and 835 Hungarian students (382 boys and 453 girls, mean age = 11.86, SD = .43) took part in the study. The sampling process emphasized background matching. The sample in China was drawn by convenience sampling. The original sample size for the Hungarian sample was much larger (N = 4790). Then, three filters (age, gender and parents' education) were set. Thus, first, Hungarian students were selected from the same age group as that of the Chinese sample. Second, the gender ratio between the two groups was equalled. Finally, third, the balance in the level of parents' education was set. The whole assessment was carried out on the eDia platform (Csapó and Molnár 2019) in the schools' ICT room. All the items were administered in simplified Chinese in China and in Hungarian in Hungary.

Instrument

The computer-based CPS test was developed in the MicroDYN approach. The test consisted of six scenarios; each scenario was covered by a unique background story. One scenario contained two items for the two phases of CPS: knowledge acquisition and knowledge application. That is, the test contained twelve items in total (six knowledge acquisition items and six knowledge application items). Item difficulty level was varied. In the MicroDYN approach, the connections between input variables and output variables can be formulated with linear structural equations (Funke 2001). Each item contains up to three input variables (see Fig. 1A, B and C) and up to three output variables (see Fig. 1X, Y and Z). Figure 1 shows several possible relations between the input and output variables. The complexity of the items was defined by the number of input and output variables and the relations between input and output variables based on the cognitive load theory (Sweller 1994). The test contained one 2–1 type (two input variables and one output variable), two 2–2 type, one 3–2 type and two 3–3 type items. The items were sorted from easy (low complexity) to difficult (high complexity).

In the first phase of a problem-solving scenario, students explored, described and operated unfamiliar systems; they sought to find out how input and output variables were interconnected; and they sought to represent this newly acquired knowledge on a concept map displayed at the bottom of the screen (knowledge acquisition; see Fig. 2; Greiff et al. 2013b). In addition, in the second phase of the problem-solving scenario, they were asked to check the system by reaching given target values (knowledge application; see Fig. 3; Greiff et al. 2013b)—the right connections could be seen on the concept map to avoid strong dependence between achievements in the first and second phases. Students had 3 min to provide answers in each phase of the problem-solving process. Figures 2 and 3 provide a screenshot of a problem with two input and output variables with two connections.

Each scenario had one fictitious cover story; thus, the contents of the problems were not based on or linked to students' real knowledge, and the relationships between the input and output variables were artificial. Students' content knowledge was therefore of no help in filling the gap between insufficient information acquired from interaction and a successful solution to the problem. Each of the input variables had five stages: + 2 (++), + 1 (+), 0, − 1 (-) and − 2 (–). To operate the system, students change the value for the input variables by clicking on the "+" or "-" button or by using the slider under the respective variable. Within the same phase of each of the problem scenarios, the history of the values for the input variables was indicated on the graph tied to each input variable. The right part of the interface contained graphs indicating the values for each of the output variables (see Fig. 4 for an example).

Each item involves four buttons: Help, Reset, Apply and Next. The problem-solver can click the Help button for a brief introduction to the operating rules of the system. If he/she needs to set the system back to its original status, that is the values for each of the input variables to zero and the values for each of the output variables to their original value, he/she can do so by clicking the Reset button. The Apply button was designed to test the effect of the currently set values for the input variables on the output variables. After clicking the Apply button, the output variable(s) will show a corresponding change in both numerical and figural

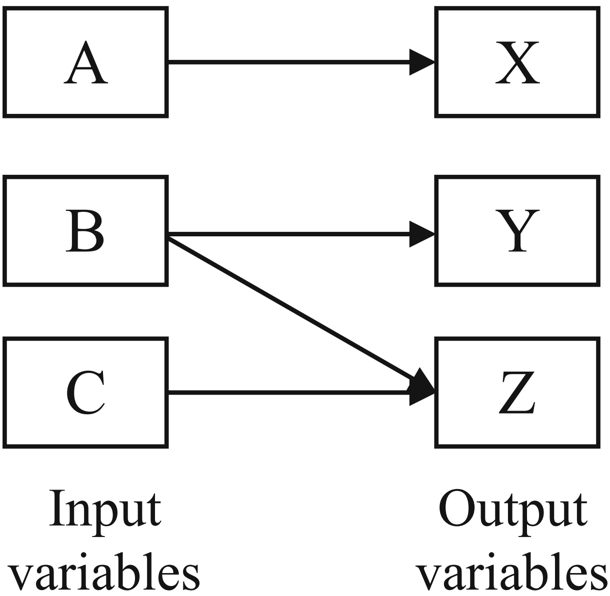

Fig. 1 A sample structure of a 3–3 type problem, with three input and three output variables with four connections

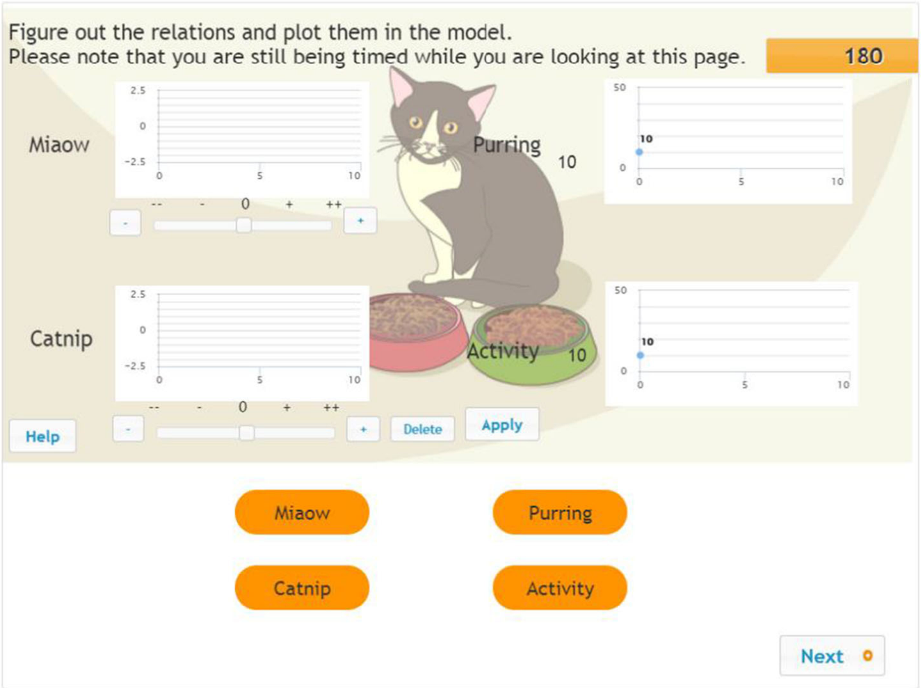

Fig. 2 Screenshot of the MicroDYN item Cat—first phase (knowledge acquisition). (The original items were in Hungarian and simplified Chinese. The controllers of the input variables range from − 2 to + 2, represented by signs ranging from "–" to "++". The current values for the output variables are displayed numerically (e.g. current value for Purring: 10) and graphically (current value: dots) (see Greiff et al. 2013b))

formats. The set of operation(s) before clicking the Apply button was defined as a trial. Within the same phase of each of the problem scenarios, the value for the input variables remained at the level of the previous input setting. The problem-solver can press the Reset button to set it back to zero or change the settings manually to any value in the available scope. He/she can also click the Next button to move to the next scenario or phase.

In the first phase, the knowledge acquisition phase, problem-solvers were asked to explore the system and find the relationships between the input and output variables. Problem-solvers were allowed to freely operate the system within 3 min. During these 3 min, they can (1) change the values for the input variables and apply trials unlimited times and (2) draw the concept map. There was no order constraint between these two activities, so the problem-solver can freely alternate between the activities. The problem-solvers need to draw the arrows between the input and output variables to present the relationships they found during their exploration (see Fig. 5 as an example). In the second phase, the knowledge application phase of the problem-solving process, students check their respective system using the concept map presented, which contains the correct connections between input and output variables. They allow the output variables to reach the given target values by changing the value for the input variables within 3 min. In this phase, they were only allowed four trials; that is they could click on the Apply button four or fewer than four times.

This study analyses students' exploration strategies in the exploration phase of the problem-solving process. Thus, when analysing the logfile data, we only focused on the first phase of

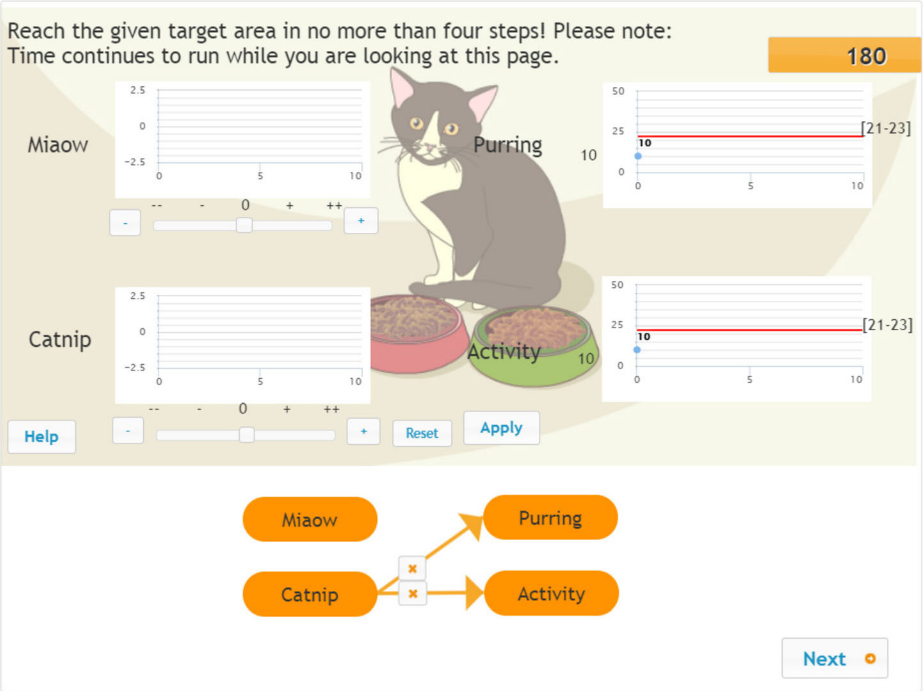

Fig. 3 Screenshot of the MicroDYN item Cat—second phase (knowledge application). (The original items were in Hungarian and simplified Chinese. The controllers of the input variables range from − 2 to + 2, represented by signs ranging from "–" to "++". The current values and the target values for the output variables are displayed numerically (e.g. current value for Activity: 10; target value: 21–23) and graphically (current value: dots; target value: red line). The correct model is shown at the bottom of the figure (see Greiff et al. 2013b))

the problem-solving process, the knowledge acquisition phase, where the problem-solver was expected to interact directly with the problem by operating the input variables, which had an influence on the output variables. We focused on the type of manipulations of the input variables and on their systematicity and effectiveness based on the amount of extracted information.

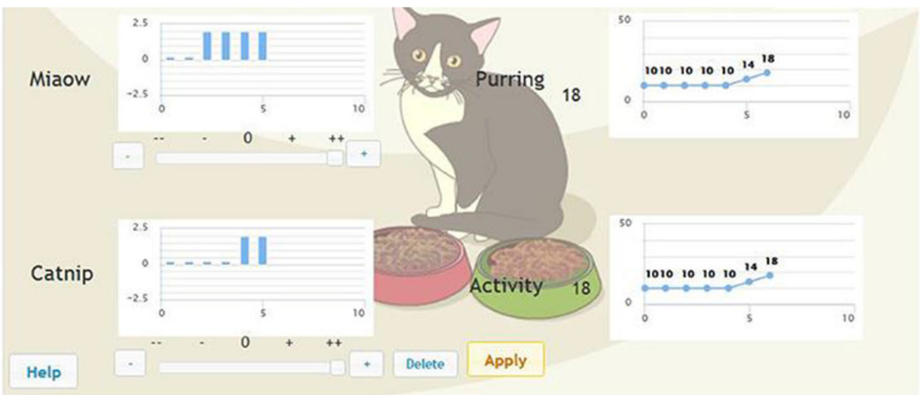

Fig. 4 Records of the changes in input and output variables

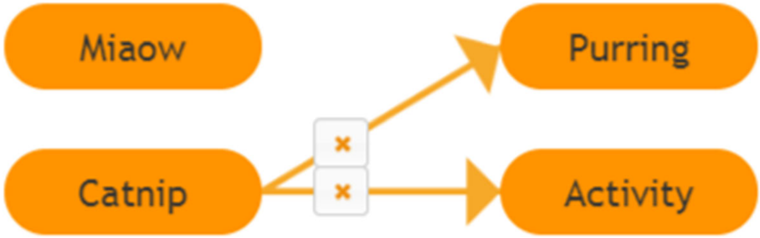

Fig. 5 Example of connecting input and output variables on a concept map

Scoring

As an achievement indicator, students' responses in the knowledge acquisition phase received a score of 1 (which means a response is correct) if the input and output variables were accurately connected on the concept map; thus, the drawings indicating the relationships between the input and output variables fully matched the underlying problem structure. Otherwise, the response received a score of 0; that is it was incorrect. For the knowledge application phase, if students successfully made the output variable reach the required range in four (or fewer) trials within the time limit, the system registered it as 1; otherwise, it was registered as 0.

Beyond the achievement data, the system logged the setting of the input variables in the knowledge acquisition phase. Each trial was formatted and labelled in the system; that is, the operation(s) the students implemented within a trial was/were summarized as one data entity in the logfile database. Students' input behaviour was defined by the sum of all trials within the same item. If an input behaviour followed meaningful regularities, it was called a strategy.

Students' minimal input behaviour model was generated from their input behaviour. It only models manipulations which were able to provide useful and new information for the problem-solvers, while some operations and activities (e.g. redundant operations and meaningless operations) were excluded. If students did not avail themselves of any useful operation based on the amount of extracted information (e.g. repetition), their strategy was marked null.

The labelling, formatting and representation method of the minimal input behaviour models employing Molnár and Csapó's (2018, p. 8) design are as follows:

1. Only one single input variable was manipulated, whose relationship to the output variables was unknown (we considered a relationship unknown if its effect cannot be known from previous settings), while the other variables were set at a neutral value like zero. We labelled this trial +1.

   Example: In the example, we assume a problem-solver is interacting with the knowledge acquisition phase of the cat problem (Fig. 2). If he/she manipulates only one input variable, for example Miaow, and another variable, for example Catnip, remains at 0, he/she can be considered to be applying a + 1 trial for the input variable Miaow.
2. One single input variable was changed, whose relationship to the output variables was unknown. The others were not at zero, but at a setting used earlier. We labelled this trial +2.

   Example: Base on the last example, the problem-solver has acquired the necessary information to determine the relationship between the input variable Miaow and the output variables since he/she applied a + 1 trial for that input variable. If he/she maintains that input variable at the setting used earlier, then manipulates the variable Catnip (the input variable whose relationship with the output variables is unknown), he/she can be considered to be applying a + 2 trial for the input variable Catnip. In this case, the problem-solver is able to determine the relationship between the input variable Catnip and the output variables using linear structural equation–based calculations.

3. One single input variable was changed, whose relationship to the output variables was unknown, and the others were not at zero; however, the effect of the other input variable(s) was known from earlier settings. Even so, this combination was not attempted earlier. We labelled this trial +3.

   Example: Back to the first example, the problem-solver has acquired the necessary information to determine the relationship between the input variable Miaow and the output variables. If he/she sets Miaow to neither 0 nor the setting used earlier and then manipulates the variable Catnip, he/she can be considered to be applying a + 3 trial for that latter input variable. In this case, he/she may still determine the relationship between that input variable and the output variables using linear structural equations-based calculations. However, the calculations will be more complex than in the case of the + 2 trial.

Moreover, the trials were modelled without the real order. The trials were sorted in ascending order in the model.

After labelling and recording students' exploration strategies in the problem-solving environment, we analysed whether students' exploration strategies were VOTAT strategies. Base on the definition of VOTAT, if a student applied a + 1, + 2 or + 3 trial for every input variable in one item, he/she can be considered to have used a VOTAT strategy in this item. Moreover, the study labelled students' exploration strategy in each item into three groups:

(1) 0 points: no VOTAT (no VOTAT was employed);
(2) 1 point: partial VOTAT (VOTAT was employed for some but not all the input variables); and
(3) 2 points: full VOTAT (VOTAT was employed for all the input variables).

Data analysis plan

A between-subject design was used to test measurement invariance by means of a multi-group confirmatory factor analysis (MGCFA). Weighted least squares and mean- and variance-adjusted (WLSMV) estimation was applied, and THETA parameterization was used with model estimation (Muthén and Muthén 2010). All the measurement models were computed with Mplus. An absolute fit index (the root mean square error of approximation (RMSEA)) and two incremental fit indices, the Tucker–Lewis Index (TLI) and the comparative fit index (CFI) (Xia and Yang 2019), were used to evaluate model fit. The measurement invariance analysis started with identifying the baseline model which fits within the overall sample and in each subgroup (Byrne and Stewart 2006). Configural invariance was tested by estimating the parameters of the baseline model in a multi-group model. It investigates whether the basic model structure is invariant across groups. Strong invariance can indicate cross-group equality in the loadings and intercepts. Finally, strict factorial invariance indicates whether the two groups have the same item residual variances (Byrne 2008). It requires cross-group equality in the loadings, intercepts and residual variances (Csapó et al. 2014). Nested model comparisons between the models were conducted using a special χ^2 difference test for the WLSMV estimator (Muthén and Muthén 2010). Measurement invariance is present if models with restrictions on model parameters (the strong and strict factorial invariance models) do not result in a significant decrease in fit relative to the unrestricted model (configural invariance model).

The independent sample t test and effect size (Cohen's d) were used to analyse the CPS performance differences between the subgroups. Students' strategy use was first represented by percentages of the application of full/partial/no VOTAT strategies in every type (defined by problem complexity) of item. χ^2 tests were used to analyse if there is any difference between Chinese and Hungarian students in their application of VOTAT. The minimal input behaviour model was used to represent students' VOTAT use afterwards. The minimal input behaviour models represented three types of VOTAT (+ 1 + 1, + 1 + 2 and + 1 + 3) for items with two input variables and three types of VOTAT (+ 1 + 1 + 1, + 1 + 1 + 2 and + 1 + 2 + 2) for items with three input variables. The study analysed the CPS performance among users of each type of VOTAT in both Chinese and Hungarian contexts, thus exploring the effectiveness of each strategy in these two cultures.

LCA (Latent Class Analysis) is a latent variable modelling approach which can be used to identify latent classes of samples who share similar observed variables (Collins and Lanza 2010). It has commonly been used to analyse explorer class profiles and behavioural pattern in CPS researches (see e.g. Mustafić et al. 2019; Molnár and Csapó 2018). It was thus employed in this study to build latent classes based on students' adoption of exploration strategies. The LCA analysis was carried out using Mplus (Muthén and Muthén 2010). The Akaike information criterion (AIC, lower values indicated a better model fit), Bayesian information criterion (BIC, lower values indicated a better model fit), adjusted Bayesian information criterion (aBIC, lower values indicated a better model fit), entropy (within a [0,1] scale, values near one, indicating high certainty in classification) and the Lo–Mendell–Rubin Adjusted Likelihood Ratio (compares the model containing n latent classes with the model containing n-1 latent classes, p value was the indicator for whether a significant difference could be detected; Lo et al. 2001) were used to approximate and decide the correct number of classes in LCA models. In addition, the average latent class probabilities (ALCP) were calculated for most likely latent class membership by latent class. This index is an indicator of the model's ability to correctly classify each individual into the most appropriate latent class profile. An ALCP above 0.8 is considered a good result (Geiser 2013).

Results

Measurement invariances were identified across gender but not nationality (RQ1)

The reliability indices were good for both the Chinese (Cronbach's $\alpha = .87$) and Hungarian (Cronbach's $\alpha = .84$) samples. The internal consistencies confirmed that the assessment was reliable. Before moving on to the next step of the analysis, an essential question needed to be answered: can CPS be measured invariantly across gender and nationality both in European and East Asian contexts? If measurement invariance is not evident, it is possible that the differences between students' ability levels are influenced by group features; that is—in the present case—they are varied due to their different cultural backgrounds and gender and not only because of the differences in their ability levels. A measurement invariance analysis was conducted across gender and nationality (see Table 1).

The results based on MGCFA (CN-HU) indicated measurement non-invariance of CPS across nationality, but measurement invariance across gender within the same culture. As the items required a minimal amount of reading and were designed as not based on and linked to students' real knowledge, we assumed that the reason for non-invariance could be found in students' use of different strategies during the problem-solving process.

Chinese students showed better problem-solving performance than Hungarian students, and a gender difference can be detected in the Hungarian sample (RQ2)

Table 2 shows the descriptive statistics for CPS and its subscales. The Chinese students achieved significantly higher scores on both the test and subtest levels with medium effect sizes. There was no gender difference ($t = 1.16$, $p > .05$, Cohen's $d = .17$) in CPS performance among Chinese males (M = 44.71%, SD = 29.03%) and Chinese females (M = 39.62%, SD = 30.34%). In contrast, the Hungarian males' CPS performance (M = 36.00%, SD = 23.32%) proved to be significantly higher ($t = 2.50$, $p < .05$) than that of the Hungarian females (M = 32.12%, SD = 21.02%). However, the effect size was small (Cohen's $d = .17$), suggesting that the detected gender difference was trivial.

The Chinese students employed VOTAT more frequently than the Hungarian students (RQ3)

VOTAT strategies were more frequently employed by the Chinese students (see Table 3). More than 56% of the Chinese students employed the full VOTAT strategy with the easiest items with two input and one or two output variables. In comparison, this percentage was 43% among the Hungarian students. For the items with three input variables (the 3–2 and 3–3 types), still more than 45% of the Chinese students employed a full VOTAT strategy, which was far higher than the rate among the Hungarian students (29–37%). These differences also manifested on the item level. According to the χ^2 tests, there were significant differences between the Chinese and Hungarian students in the application of VOTAT on every type of problem that could be identified. To sum up, generally, the 12-year-old Chinese and Hungarian students employed different exploration strategies during the problem-solving process. The Chinese students were more successful in the application of the best, most informative strategy.

Table 1 Goodness of fit indices for testing invariance across nationality and gender of CPS

Group	Invariance model	$\chi2$	df	p	CFI	TLI	RMSEA	$\Delta\chi2$	p
Nationality	Config. inv.	136.88	65	< .01	.99	.99	.05	–	–
	Strong factorial inv.	307.73	69	< .01	.95	.96	.08	267.00	< .01
	Strict factorial inv.	353.66	64	< .01	.94	.95	.09	187.76	< .01
Gender (HU)	Config. inv.	88.74	63	< .05	.99	.99	.03	–	–
	Strong factorial inv.	90.74	67	< .05	.99	.99	.03	6.30	.50
	Strict factorial inv.	88.42	66	< .05	.99	.99	.03	16.17	.30
Gender (CN)	Config. inv.	65.35	43	< .05	.99	.99	.08	–	–
	Strong factorial inv.	66.78	46	< .05	.99	.99	.07	3.43	.75
	Strict factorial inv.	54.57	45	.16	1.00	1.00	.05	4.54	.95

Table 2 Chinese and Hungarian students' performance on CPS test and its subscales

Test	Nationality	Mean (%)	SD (%)	t	*p*	Cohen's *d*
Knowledge acquisition	CN	47.86	34.19	2.94	< .01	.25
	HU	39.92	29.32			
Knowledge application	CN	36.01	31.04	3.41	< .01	.31
	HU	27.86	21.39			
Complex problem-solving	CN	41.93	29.78	3.48	< .01	.31
	HU	33.89	22.17			

Using VOTAT had more value for the Chinese students in a CPS environment (RQ4)

Results (see Figs. 6 and 7) indicated that use of the strategy labelled + 1 + 1 (or + 1 + 1 + 1 for items with three input variables) resulted in the right solution to the problem, independent of the complexity of the problem, with the highest probability in both cultures; thus, this strategy proved to be the most effective one in the complex problem-solving environment built in this study. In the + 1 + 1 (or + 1 + 1 + 1) type strategy, for every operation, only one single input variable was manipulated, while the other variables were set at a neutral value (at zero in the case of MicroDYN). This is a typical VOTAT strategy, which was called *VOTAT strategy A* in this study.

The second most frequently employed VOTAT strategy was the + 1 + 2 type (in the case of two input variables) or the + 1 + 2 + 2 type (in the case of three input variables) strategy. This strategy was named *VOTAT strategy B*. In this type of strategy, in the first step, only one single input variable was manipulated, while the other variables remained at zero; in the second step, the input variable which was changed in the first step remained at the previous status, while the other input variable was manipulated. This type of strategy also provides a certain probability of correctly solving the problem. However, the users of this type of strategy also had a comparatively high chance of supplying a wrong answer. In most cases, the chance of arriving at a wrong answer was higher or almost equal to the possibility of furnishing a right answer (except for items with two input variables and two output variables in the Chinese context). Thus, the effectiveness of this kind of strategy was not very high, especially compared with *VOTAT strategy A*.

The + 1 + 3 type (in the case of two input variables) and the + 1 + 1 + 2 type (in the case of three input variables) were rarely employed by the participants. Thus, their effectiveness was difficult to accurately determine. In this case, these strategies will not be discussed any further.

In comparing the exploration behaviour of the Chinese and Hungarian students, we identified some remarkable differences:

Table 3 Frequencies of VOTAT strategy usage for Chinese and Hungarian students

Complexity	No VOTAT (%)		Partial VOTAT (%)		Full VOTAT (%)		χ^2	*p*
	CN	HU	CN	HU	CN	HU		
2–1	33.7	42.6	4.8	14.0	61.5	43.4	24.25	< .01
2–2	38.1	44.9	5.9	11.6	56.2	43.7	18.48	< .01
3–2	41.7	58.4	5.9	12.0	52.4	29.7	27.74	< .01
3–3	47.9	53.9	7.0	9.7	45.2	36.5	14.81	< .01

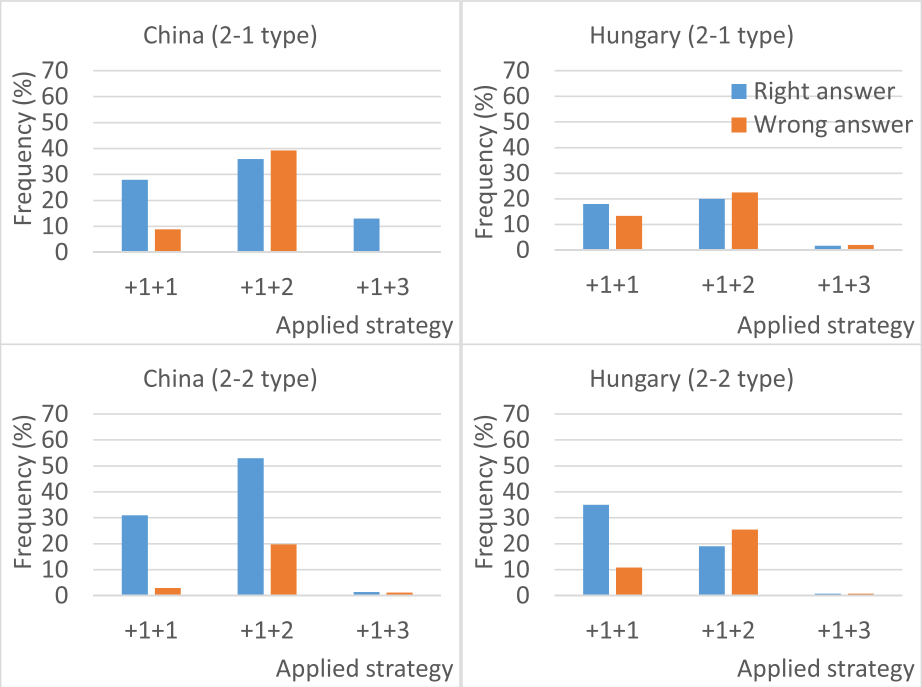

Fig. 6 Efficacy of the most frequently employed VOTAT strategies on items with two input variables

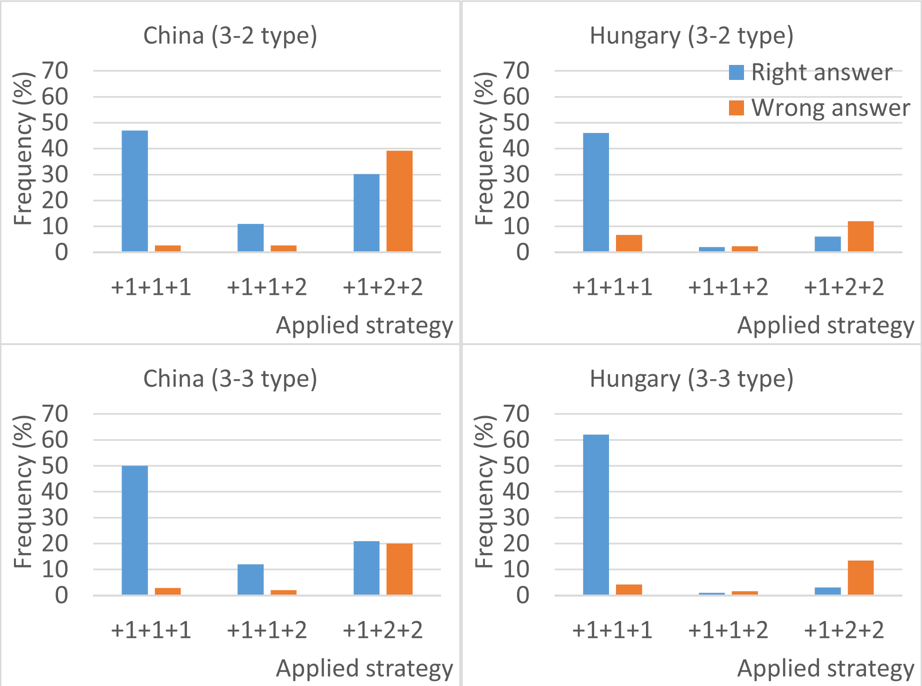

Fig. 7 Efficacy of the most frequently employed VOTAT strategies on items with three input variables

(1) *VOTAT strategy A* has been proved as the most effective strategy. That is, the *VOTAT strategy A* users had a high probability of arriving at the correct answer. However, in every item, the proportions for Hungarian *VOTAT strategy A* users with a wrong answer were higher than that of the Chinese users. All the *VOTAT strategy A* users extracted the same information from the system. Thus, the ability to effectively use the extracted information was the decisive factor in whether they could successfully solve the problem. Therefore, compared to Chinese students, Hungarian students had a lower ability to effectively use the extracted information.

(2) The frequency for Chinese students employing *VOTAT strategy B* was higher than the same frequency for the Hungarian students. *VOTAT strategy B* had higher effectiveness in the Chinese context. The result confirmed Chinese students' higher ability level to effectively use the extracted information. Moreover, it showed an effective strategy use was more valuable for the Chinese problem-solvers in a CPS environment, since they could use the extracted information more effectively.

(3) When comparing *VOTAT strategy A* with *VOTAT strategy B*, the latter is more complex. If a student employed *VOTAT strategy B*, he/she needed to do a more complex analysis and calculation, which required combinatorial reasoning skill and mathematical computation ability. Students have the option of employing a VOTAT strategy with accidental random operations. Of course, if a student employed a VOTAT strategy involuntarily, but did not fully understand what he/she had done, he/she was not very likely to effectively use the extracted information and solve the problem correctly. As the results indicated (Figs. 6 and 7), a significantly higher proportion of Chinese students employed *VOTAT strategy B* and they had a much higher chance of arriving at the correct answer by employing that strategy. That is, there were far more Chinese students who understood how VOTAT worked in a CPS environment during the assessment process.

Three qualitatively different explorer class profiles can be distinguished in both the Chinese and Hungarian samples; the Chinese and Hungarian students showed different behavioural patterns (RQ5)

Latent class analyses were conducted based on students' minimal input behaviour. The 3-latent class model had the highest entropy and comparatively low information theory criteria fits (AIC, BIC and aBIC) in the Chinese sample. The likelihood ratio statistical test (Lo–Mendell–Rubin adjusted likelihood ratio test) confirmed that the 3-latent class model is the best fitting model solution (Table 4). In the Hungarian context, the 3-latent class model had relatively low information theory criteria fits; however, its entropy was lower than that of the 2-latent class

Table 4 Fit indices for latent class analyses

Nationality	Number of latent classes	AIC	BIC	aBIC	Entropy	L–M–R test	*p*
CN	2	1475	1556	1477	0.968	504	< .01
	3	1461	1584	1464	0.977	40	< .05
	4	1458	1622	1461	0.973	29	> .05
HU	2	7791	7909	7830	0.930	1873	< .01
	3	7550	7730	7609	0.817	264	< .01
	4	7511	7752	7590	0.787	64	> .05

model. The likelihood ratio statistical test proved that the 3-latent class model fitted the data better ($p < .01$). Thus, both the Chinese and Hungarian samples were classified for three latent classes.

The three qualitatively different class profiles for both the Chinese and Hungarian samples are shown in Figs. 8 and 9. 45.2% of the Chinese students fall within a group of students who were intermediate performers on the easiest items, but low performers on the complex ones with a very slow learning effect (Class 1); 3.1% were rapid learners (Class 2); and finally 51.7% proved to be proficient strategy users (Class 3; Fig. 8). The ALCPs were .99, .95 and .99 for these three latent classes, respectively. Meanwhile, 28.4% of the Hungarian students are among the Class 1 students, who employed a full or partial VOTAT strategy on the easiest items, but did not learn to check and understand the strategy. After problem complexity increased, they became low performers. 34.1% proved to be proficient strategy users (Class 3), who employed a VOTAT strategy from the beginning to the end, and 37.5% were low performers (Class 4), who rarely employed a full or partial VOTAT strategy. The ALCPs were .85, .97 and .91 for the three latent classes in the Hungarian samples, respectively. The high ALCP for every latent class proved the LCA models have the ability to correctly classify each individual into the most appropriate latent class profile.

There were several remarkable differences identified in the comparison of Chinese and Hungarian students' VOTAT strategy use. The class of "low performers" did not exist in the Chinese sample, indicating that there were almost no Chinese students who did not employ the VOTAT strategy at least once in the CPS test. 45.2% of the Chinese students were intermediate performers on the easiest items, but low performers on complex ones with a very slow learning effect. They did not perform well on the whole test, but at least they showed an average performance on the easy items.

The proportion of proficient strategy users in the Chinese sample was 51.7%, which was much higher than the percentage for the Hungarian sample (34.1%). This was an obvious and direct indicator that Chinese students performed generally better as regards VOTAT strategy learning, understanding and use. A small group of Chinese students (3.1%; rapid learners) did not employ a full or partial VOTAT strategy on the easy items at the beginning of the test, but they learned very quickly and reached the top performers' proficiency level by the end of the test. No rapid learner was found in the Hungarian sample, thus contradicting Molnár and Csapó's (2018) finding of six qualitatively different explorer class profiles in a large-scale empirical study, including a class of rapid learners (4.4% of the whole sample). The reason for this may be the large differences in the sample size of the two studies.

Discussions and limitations

The paper has focused on 12-year-old Chinese and Hungarian students' exploration behaviour and problem-solving skills in a MicroDYN-based complex problem-solving environment analysing process data. Results indicate measurement invariance across gender but not nationality. There barely exists any research on whether measurement invariance obtains using basically the same complex problem-solving test across East Asian and European assessments, especially between China and Hungary. Even though both of the countries participated in the problem-solving module of the PISA 2012 assessment, measurement invariance was not tested (OECD 2014). The MicroDYN-based CPS assessment has been shown to be measurement invariant across nationality and gender in a Hungarian and German cross-national study

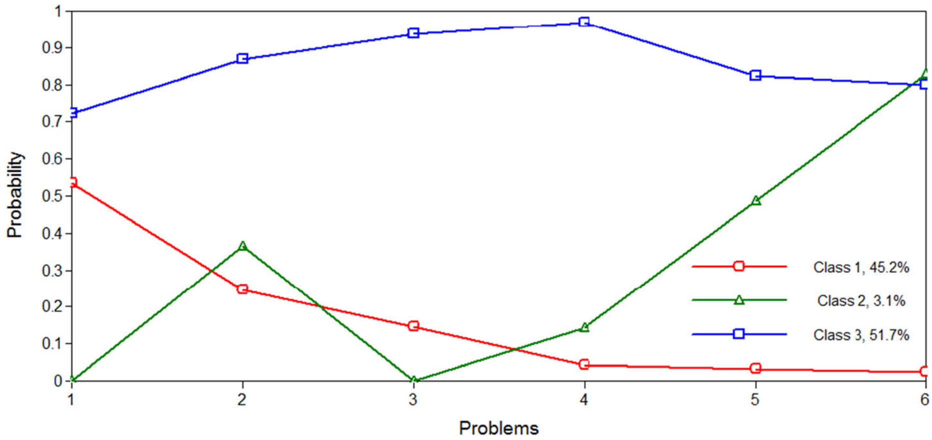

Fig. 8 Three qualitatively different class profiles for the Chinese sample

(Wüstenberg et al. 2014b). However, Wüstenberg et al. (2014b) drew their samples from two European countries with close educational and cultural links. The educational traditions and cultural differences between Hungary and Germany are far smaller than those between Hungary and China. Wüstenberg et al.'s (2014b) result is evidence that students socialized in similar cultures and school systems understand and interpret CPS problems almost the same way, showing similar behavioural patterns during the problem-solving process. Meanwhile, the present paper has indicated that students from different educational traditions and cultures might behave remarkably differently in the same problem-solving environment.

Fifteen-year-old Chinese students showed better performance than in CPS assessment (OECD 2014) in PISA 2012. Our results confirmed the remarkable gap between Chinese and Hungarian students' problem-solving achievement even earlier in the problem-solving developmental scale among 12-year-old students. That is, the large achievement differences identified among 15-year-old students are already present 3 years earlier. Thus, the cause for the differences indicated by PISA goes back much earlier. Our study explored the reasons for the different CPS performance between Chinese and Hungarian students from the perspective of process data by analysing the behavioural patterns of the students. Results indicated the

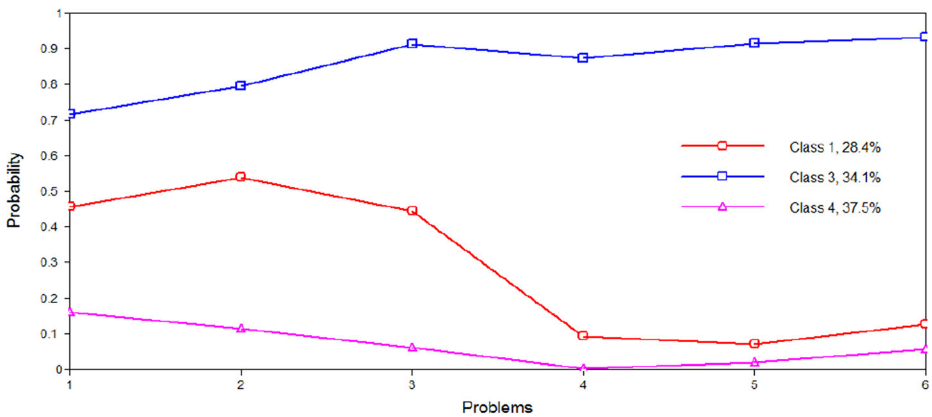

Fig. 9 Three qualitatively different class profiles for the Hungarian sample

Chinese students employed the VOTAT strategy more frequently than the Hungarian students—which confirms analyses conducted by Greiff et al. (2015c) on the PISA data—and they showed a significantly higher learning effect than their Hungarian peers. Compared with the Hungarian students, the Chinese students managed to represent the information they obtained from the system more effectively, and they made good decisions in the problem-solving process. That is, employing the VOTAT strategy had more value for the Chinese students, since if they collected enough information to solve the problems, they had higher chances of arriving at the right solution. In contrast, Hungarian students' problem-solving performance received less influence from the exploration strategies they employed. However, the Hungarian students proved to be more intuitive, and, as a result, the Hungarian theoretically non-effective strategy users had a higher probability of correctly solving the problem (compared to the Chinese students).

There is reason to believe the Hungarian students' style of strategy use will have a larger negative impact on their CPS achievement when they face a more complex situation. According to Molnár and Csapó's (2018) findings, the higher the complexity and difficulty of a problem, the lower the possibility of correctly solving it without effective strategy use. Therefore, Chinese students' higher proficiency level in effective strategy use could play a very important role in the CPS environment with higher complexity and difficulty, but this will not be the case for the Hungarian students. Thus, Hungarian students' ineffective strategy use is very likely one of the major reasons why they showed unsatisfactory performance on some CPS assessments, which were designed for older students and contained more complex items.

To sum up, the study effectively answered the research questions, which contributed to our further understanding of the nature of problem-solving in European and East Asian contexts. The results shed new light on the role of exploration strategies in the learning processes, especially in complex problem-solving environments. The findings also have direct implications for devising instructional methods to improve students' CPS skills by learning about their own learning strategies and their effectiveness. The Hungarian education system mainly focuses on the disciplinary dimension of knowledge—as can also be seen in the OECD PISA results (see OECD 2020)—and less attention has been paid to developing students' application dimension of knowledge and their thinking skills. One of the reasons could be the lack of easy-to-use assessments, and thus the cognitive development and reasoning dimension of learning remains hidden (see Molnár and Csapó 2019). Previous studies have indicated that children's reasoning skills can be developed effectively (Klauer and Phye 2008; Molnár 2011; Perret 2015); that is one of the solutions to the problem indicated above could be that the development of thinking skills should become an integral part of the school curriculum. In addition, it should play a part in a wide variety of school learning activities.

The results and findings provide a sound basis for future studies. A large number of studies have focused on problem-solving skills from different perspectives. However, the component skills of problem-solving, especially in international context, and the role of reasoning skills (e.g. inductive and combinatorial reasoning, see Wu and Molnár 2018) in successful problem-solving or the influential power of everyday learning strategies on a student's developmental level of problem-solving skills still call for further study. Through logfile analysis and process data, we have the chance to acquire a deeper understanding of students' thinking, problem-solving, exploration and learning strategies and of the reasons behind their achievement, e.g. to differentiate students with similar achievement, but using different strategies (see e.g. Molnár and Csapó 2018).The results indicate the possibility for further research to discover a larger picture of the connections between students' cognitive features and physical behaviour in the problem-solving environment. The remarkable difference between Chinese and Hungarian students' strategy use can be further explored and explained from a cognitive perspective.

There are some limitations in this cross-national comparison study. The measurement non-invariance detected across nationalities, on the one hand, is a very interesting and important finding; on the other hand, it has the potential to impact the accuracy of the results, especially as regards the cross-national comparison. Currently, we assume the reason for non-invariance may lie in Chinese and Hungarian students' different cognitive styles and behaviour during the CPS progress. Future studies are called for to explore the reasons. The relatively small Chinese sample size caused some limitations regarding the validity of the findings, such as the small size for the rapid learner class profile in the Chinese sample, which thus did not yield a particularly valid result. The sample size for the assessment differed in China and Hungary. However, to increase the reliability and validity of the results, a perfect match was made; that is the Hungary sample was perfectly matched to the Chinese sample based on students' age, gender and level of parents' education, thus excluding background factors which proved to have a significant effect on achievement differences based on the literature (Csapó and Molnár 2017; Molnár et al. 2013; Wüstenberg et al. 2014b; OECD 2014; Wittmann and Hattrup 2004). Based on the current small-scale results, some interesting trends have been noted, which may form a solid foundation for the next empirical studies, focusing on students' exploration behaviour in a CPS environment. Furthermore, as the current results are based on the background matched samples, we put forward the hypothesis that larger performance and behaviour differences in line with the current results will be detected in a large-scale study (which is also indicated between Hungary and China in PISA). Further large-scale studies are required to validate the results. Last but not least, considering that the Chinese sample was small and that the Hungarian sample was a subsample of a larger and more representative sample, our samples were neither representative for the Chinese students nor for their Hungarian peers. This may also cause concern about generalizability for East Asian and European countries. Future studies will continue to include more countries to discover such differences in the cognitive structures for problem-solving that exist between students with East Asian and European cultural backgrounds.

Funding This study was funded by EFOP-3.4.3-16-2016-00014.

Data availability The data used to support the findings cannot be shared at this time as the data also forms part of an ongoing study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Code availability The codes used to analyse the data are available from the corresponding author upon reasonable request.

References

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 27–63). NJ: Erlbaum, Hillsdale.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema, 20*(4), 872–882.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: a walk through the process. *Structural Equation Modeling, 13*(2), 287–321.

Chen, Z., & Klahr, D. (1999). All other things being equal: acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Chen, Y., Li, X., Liu, J., & Ying, Z. (2019). Statistical analysis of complex problem-solving process data: an event history analysis approach. *Frontiers in Psychology, 10*, 486.

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis. In *With applications in the social, behavioral, and health sciences*. New York: Wiley.

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012). Technological issues of computer-based assessment of 21st century skills. In B. McGaw, P. Griffin, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 143–230). New York, NY: Springer.

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology, 8*, 2022.

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia System. *Frontiers in Psychology, 10*, 1522.

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

Dörner, D., & Funke, J. (2017). Complex problem solving: what it is and what it is not. *Frontiers in Psychology, 8*, 1153.

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Unigang mit Unbestimmtheit und Komplexität*. Bern: Huber.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving, 4*(1), 19–42.

Frensch, P. A., & Funke, J. (1995). *Complex problem solving: The European perspective*. New York: Psychology Press.

Funke, J. (1991). Solving complex problems: exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). New York: Psychology Press.

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

Funke, J., & Frensch, P. A. (2007). Complex problem solving: the European perspective - 10 years after. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 25–47). New York: Lawrence Erlbaum.

Geiser, C. (2013). *Data analysis with Mplus*. New York, NY: The Guilford Press.

Greiff, S., & Funke, J. (2009). Measuring complex problem solving-the MicroDYN approach. In F. Scheuermann (Ed.), *The transition to computer-based assessment-lessons learned from largescale surveys and implications for testing* (pp. 157–163). Luxembourg: Office for Official Publications of the European Communities.

Greiff, S., & Wüstenberg, S. (2014). Assessment with microworlds: factor structure, invariance, and latent mean comparison of the MicroDYN test. *European Journal of Psychological Assessment, 30*, 1–11.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: a new measurement perspective. *Applied Psychological Measurement, 36*(3), 189–213.

Greiff, S., Holt, D., & Funke, J. (2013a). Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *Journal of Problem Solving (The), 5*, 71–91.

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

Greiff, S., Kretzschmar, A., Müller, J., Spinath, B., & Martin, R. (2014a). Computer-based assessment of complex problem solving in educational contexts and how it is influenced by students' level of information and communication technology literacy. *Journal of Educational Psychology, 106*(3), 666–680.

Greiff, S., Wusterberg, S., Csapo, B., Demetriou, A., Hautamaki, J., Graesser, A., & Martin, R. (2014b). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13*, 74–83.

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

Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015b). Sometimes less is more: comparing the validity of complex problem solving measures. *Intelligence, 50*, 100–113.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015c). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education, 91*, 92–105.

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.

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: a training approach. *Review of Educational Research, 78*(1), 85–123.

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 371–393). Oxford: Wiley-Blackwell.

Kuhn, D. (2012). The development of causal reasoning. *WIREs Cgnitive Science, 3*(3), 327–335.

Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*, 1–157.

Leutner, D., Wirth, J., Klieme, E., & Funke, J. (2005). Ansätze zur Operationalisierung und deren Erprobung im Feldtest zu PISA 2000. In E. Klieme, D. Leutner, & J. Wirth (Eds.), *Problemlösekompetenz von Schülerinnen und Schülern* (pp. 21–36). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lien, Y. C. N., Wu, W. J., & Lu, Y. L. (2020). How well do teachers predict students' actions in solving an ill-defined problem in STEM education: a solution using sequential pattern mining. *IEEE Access, 8*, 134976–134986.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 287–303). Mahwah, NJ: Lawrence Erlbaum.

Molnár, G. (2011). Playful fostering of 6-to 8-year-old students' inductive reasoning. *Thinking skills and Creativity, 6*(2), 91–99.

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

Molnár, G., & Csapó, B. (2019). Making the psychological dimension of learning visible: using technology-based assessment to monitor students' cognitive development. *Frontiers in Psychology, 10*, 1368.

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9*, 35–45.

Molnár, G., Greiff, S., Wüstenberg, S., & Fischer, A. (2017). Empirical study of computer-based assessment of domain-general dynamic problem solving skills. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 123–143). Paris: OECD Publishing.

Mustafić, M., Yu, J., Stadler, M., Vainikainen, M. P., Bornstein, M. H., Putnick, D. L., & Greiff, S. (2019). Complex problem solving: profiles and developmental paths revealed via latent transition analysis. *Developmental Psychology, 55*(10), 2090–2101.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321–349). Cambridge: University Press.

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

OECD. (2020). *PISA 2018 results (volume I): what students know and can do*. Paris: OECD Publishing.

Perret, P. (2015). Children's inductive reasoning: developmental and educational perspectives. *Journal of Cognitive Education and Psychology, 14*(3), 389–408.

Ren, Y., Luo, F., Ren, P., Bai, D., Li, X., & Liu, H. (2019). Exploring multiple goals balancing in complex problem solving based on log data. *Frontiers in Psychology, 10*, 1975.

Schoppek, W., & Fischer, A. (2017). Common process demands of two complex dynamic control tasks: transfer is mediated by comprehensive strategies. *Frontiers in Psychology, 8*, 2145.

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.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: new possibilities for understanding students' problem-solving skills. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: using research to inspire 21st century learning* (pp. 193–209). Paris: OECD Publishing.

UNESCO Institute for Statistics. (2009). *Guide to measuring information and communication technologies (ICT) in education*. Montreal: UNESCO Institute for Statistics.

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14*(4), 541–554.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*(1), 75–100.
Wenke, D., Frensch, P. A., & Funke, J. (2005). Complex problem solving and intelligence: empirical relation and causal direction. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 160–187). New York: Cambridge University Press.
Wittmann, W. W., & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research, 21*(4), 393–409.
Wu, H., & Molnár, G. (2018). Interactive problem solving: assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research, 26*(1), 90–105.
Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving – more than reasoning? *Intelligence, 40*(1), 1–14.
Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014a). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29*, 18–29.
Wüstenberg, S., Stadler, M., Hautamäki, J., & Greiff, S. (2014b). The role of strategy knowledge for the application of strategies in complex problem solving tasks. *Technology, Knowledge and Learning, 19*(1–2), 127–146.
Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behavior Research Methods, 51*(1), 409–428.
Zoanetti, N., & Griffin, P. (2017). Log-file data as indicators for problem-solving processes. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 177–191). Paris: OECD Publishing.

Hao Wu. Research Group on Learning and Instruction, University of Szeged, Szeged, Hungary

Current themes of research:

ICT in education. Cognitive skills development. Technology-based assessment. E-learning

Most relevant publication in the field of Psychology of Education:

Wu, H., & Molnár, G. (2018). Interactive problem solving: assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research, 26*(1), 90–105.

Gyöngyvér Molnár. Department of Learning and Instruction, MTA–SZTE Research Group on the Development of Competencies, University of Szeged, Szeged, Hungary

Current themes of research:

Technology-based assessment. ICT in education. Measuring and enhancing reasoning skills

Most relevant publications in the field of Psychology of Education:

Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of educational psychology, 106*(3), 639.
Molnár, G. (2011). Playful fostering of 6-to 8-year-old students' inductive reasoning. *Thinking skills and Creativity, 6*(2), 91–99.
Molnár, G., & Csapó, B. (2018). The efficacy and development of students' problem-solving strategies during compulsory schooling: Logfile analyses. *Frontiers in Psychology, 9*, 302.
Molnár, G., & Csapó, B. (2019). Making the psychological dimension of learning visible: using technology-based assessment to monitor students' cognitive development. *Frontiers in psychology, 10*, 1368.
Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity, 9*, 35–45.

Molnár, G., Greiff, S., Wüstenberg, S. & Fischer, A. (2017). Empirical study of computer-based assessment of domain-general dynamic problem solving skills. In B. Csapó and J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 123–143). Paris: OECD Publishing.