

# Exploring Exploration: Inquiries into Exploration Behavior in Complex Problem Solving Assessment

Jonas C. Müller  
jonas.mueller@uni.lu

André Kretschmar  
andre.kretschmar@uni.lu

Samuel Greiff  
samuel.greiff@uni.lu

University of Luxembourg  
6, rue Richard Coudenhove Kalergi  
1359 Luxembourg-Kirchberg

## ABSTRACT

Complex Problem Solving (CPS) is a prominent representative of transversal, domain-general skills, empirically connected to a broad range of outcomes and recently included in large-scale assessments such as PISA 2012. Advancements in the assessment of CPS are now calling for a) broader assessment vehicles allowing the whole breadth of the concept to unfold and b) additional efforts with regard to the exploitation of log-file data available. Our Paper explores the consequences of heterogeneous tasks with regard to the applicability of an established measure of strategic behavior (VOTAT) featured currently in assessment instruments. We present a modified conception of this strategy suitable for a broader range of tasks and test its utility on an empirical basis. Additional value is investigated along the line of theory driven educational data mining of process data.

## Keywords

Complex Problem Solving, Theory Driven Data Mining, Exploration Behavior.

## 1. INTRODUCTION

Targeting human behavior in problem situations characterized by dynamic and interactive features [1], widespread application of Complex Problem Solving (CPS) assessment only began after the introduction of formal frameworks and the restriction to so-called minimal complex systems [3]. Recent inclusions of CPS as a representative of domain-general skills in large-scale studies such as the Programme for International Student Assessment (PISA) in 2012 can be seen as a direct result of these advancements.

### 1.1 CPS assessment with finite-state automata

Initially allowing for the reliable assessment of CPS, thereby building towards the foundation of the success of the concept, the restriction to tasks based on the formal framework of Linear Structural Equations (LSE) is at the same time restrictive with regard to the kind of problems that can be modeled: Only *quantitative* relations between variables can be simulated in assessment. This restriction can be hindering in several ways as it is limiting the necessary behavior for successful handling of the problem. Therefore, current work in the domain [3, 4] is targeting

the inclusion of tasks based on a second formal framework: Finite State Automata (FSA), thereby (re-)introducing a whole range of problem features to task construction and assessment [2].

### 1.2 Strategies in CPS assessment

The availability of process data documenting participant's behavior while working on a CPS task presents an ideal point of departure for educational data mining along the line of Sao Pedro et al. [5]. The problem lies with defining meaningful patterns, that can subsequently be used in text replay tagging [5]: Basic analysis based on behavioral measures like the number of interactions or time working on a task has been rather disappointing from the view of explaining variations in CPS performance. Fortunately enough, a strategy called vary-one-thing-at-a-time (VOTAT, also called control-of-variables-strategy, CVS) can be considered optimal for exploring LSE-based tasks [3]. Process analysis built on this strategy has been widely employed in CPS assessment.

### 1.3 Adapting VOTAT to FSA-based tasks

The introduction of FSA-based tasks, however, is making the analysis and scoring of VOTAT-behavior insufficient. In some cases, a participant solely applying VOTAT will not even reach essential states for the task's understanding. To account for this change in optimal strategic behavior, we adapted the search for adequate behavior to the possibilities of FSA-based tasks.

Tasks based on LSE show underlying relations that stay the same during the whole course of working on them: The effects of increasing variable A will always result in a reduction of the value in variable Y, making VOTAT the optimal exploration strategy. In FSA-based tasks on the other hand, the effect of increasing an input variable might not only depend on that variable, but also on the input values of other variables, specific arrangements of inputs (e.g., one high, one low), or prior output variable values.

As a result, adequate explorative behavior has to take the state into account the task is currently in: A resulting behavioral pattern, we call nested-VOTAT, is referring to different 'areas' of the system in which input variations are showing qualitatively comparable effects. Nested-VOTAT is shown, when isolated input variations (i.e., a VOTAT-like behavior) are used in these areas: Systematically exploring the effects of input variation in isolation and for all input variables within an area of similar task functioning. This scoring can subsequently be used in procedures like text replay tagging [5].

### 1.4 Research question

We are interested in the applicability and empirical usefulness of this extension of VOTAT for FSA-based tasks, especially when empirically related to classical tasks based on LSE. The approach is mixing exploratory elements of data analysis with theoretically driven assumptions.

## 2. METHOD

**Participants and procedure:** Participants provided data in the context of a broader assessment at a German school in November 2012. There were 565 students who completed the relevant assessment of CPS, 54% of participants were female, mean age was 14.95 years ( $SD = 1.30$ ).

**Measures:** *FSA:* Five LSA-based tasks were included in the assessment. The test is called MicroFIN [4], ‘Micro’ standing for the minimal complex systems approach and ‘FIN’ for the framework of finite state automata. Separate phases of MicroFIN are targeting knowledge acquisition and knowledge application. Empirically, we expect a three-dimensional measurement model for MicroFIN, separating the application of nested-VOTAT, knowledge acquisition, and knowledge application.

*Scoring of nested-VOTAT:* Credit for nested-VOTAT is given per task for the isolated variation of all input variables in one of the qualitatively different areas of the task. The number of these areas varies with the number of states the problem shows different relations to input variations for.

*LSE:* To get a comparable measure of CPS based on LSE, eight MicroDYN tasks were included in testing. MicroDYN [3] includes a scoring of VOTAT in the exploration phase, which is typically strongly connected to knowledge acquisition. A two-dimensional measurement model is expected, conflating strategy application (i.e., VOTAT) and knowledge acquisition [3].

## 3. RESULTS

**Manifest Correlations:** Manifest correlations between nested-VOTAT indicators and knowledge acquisition and knowledge application in MicroFIN were of rather low size per task. This result is in line with previous findings with regard to single task indicators of MicroFIN and MicroDYN.

**Dimensionality and internal consistency:** Exploratory factor analysis for nested-VOTAT indicated a single latent factor. Internal consistency was poor with Cronbach’s alpha  $\alpha = .53$ .

**Latent modeling:** *Measurement models:* Measurement models for MicroFIN indicated the separability of nested-VOTAT application and knowledge acquisition, as well as a third dimension for knowledge application. The three dimensional model fitted well, conflating any of the two dimensions resulted in significantly worse model fit ( $\chi^2(167) = 230.246, p < .001$ , RMSEA = 0.026, CFI = 0.981, TLI = 0.978; latent factors correlated  $r = .71$  to  $.81$ , all  $p < .001$ ). Measurement models for MicroDYN indicated the expected two-dimensional model fitting well ( $\chi^2(251) = 588.234, p < 0.001$ , RMSEA = 0.049, CFI = 0.994, TLI = 0.993, latent correlation of factors  $r = .79, p < .001$ ). Separating the use of VOTAT from knowledge acquisition in MicroDYN resulted in estimation problems due to very highly correlated factors.

*Structural models:* Relations between latent indicators for MicroFIN and MicroDYN can be found in Table 1.

Latent correlations between MicroFIN and MicroDYN ( $r = .56$  to  $.68$ ) were slightly lower than the internal correlations between facets within each instrument ( $r = .71$  to  $.81$ ). No major difference with regard to the facets could be found: Knowledge acquisition facets were not stronger related to each other, than to the knowledge application facet of the other instrument. The facet indicating the application of nested-VOTAT in MicroFIN did show a stronger connection to knowledge acquisition, than knowledge application in MicroFIN ( $p < .001$ ).

**Table 1: Latent correlations of MicroFIN and MicroDYN**

Test	Facet	MicroFIN		MicroDYN	
		(1)	(2)	(3)	(4)
MicroFIN	(2)	0.81			
	(3)	0.80	0.71		
MicroDYN	(4)	0.66	0.64	0.61	
	(5)	0.68	0.56	0.62	0.78

*Note:* (1) Knowledge acquisition in MicroFIN, (2) Knowledge application in MicroFIN, (3) Use of nested-VOTAT in MicroFIN, (4) Knowledge acquisition in MicroDYN (incl. use of VOTAT), (5) Knowledge application in MicroDYN. All correlations are significant on a .01 level.

## 4. DISCUSSION

The present study represents a first advance into utilizing the potential of process data in FSA-based tasks. It shows the feasibility of including measures of exploration behavior into the assessment of CPS even when based on heterogeneous tasks.

The notion of nested-VOTAT has been shown to be applicable to a range of tasks based on FSA with a varying degree of success: The low internal consistency of  $\alpha = .53$  could mean different optimal strategies being necessary in some of the tasks, as it could be related to measurement problems due to the high heterogeneity of tasks. A broader variation of item features and alternative strategy scorings is needed to clarify this aspect.

The latent correlations point to the potential of including process-related measures into FSA-based CPS assessment: Contrary to the case of LSE-based tests, the higher heterogeneity of FSA seems to result in a more probabilistic relation between strategies in exploration and knowledge acquisition.

Future inquiries into the determinants of CPS performance should be built on broad assessment vehicles and careful examination of participant behavior. Based on the findings elaborated here, approaches of educational data mining like text replay tagging can be utilized even when targeting a heterogeneous set of tasks.

## 5. REFERENCES

- [1] Buchner, A. 1995. Basic topics and approaches to the study of complex problem solving. *Complex problem solving: The European perspective*. P.A. Frensch and J. Funke, Eds. 27–63.
- [2] Buchner, A. and Funke, J. 1993. Finite-state automata: Dynamic task environments in problem-solving research. *The Quarterly Journal of Experimental Psychology Section A*. 46, 1 (1993), 83–118.
- [3] Greiff, S., Wüstenberg, S. and Funke, J. 2012. Dynamic Problem Solving: A new assessment perspective. *Applied Psychological Measurement*. 36, 3 (2012).
- [4] Müller, J.C., Kretzschmar, A., Wüstenberg, S. and Greiff, S. 2013. Extending the Assessment of Complex Problem Solving to Finite State Automata: The Merits of Heterogeneity. Manuscript submitted for publication.
- [5] Sao Pedro, M.A., Baker, R.S.J.d., Gobert, J.D., Montalvo, O. and Nakama, A. 2013. Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction*. 23, 1 (2013), 1–39.