An Evaluation of Elementary Students' Ability of Problem Solving in Information Processing

Lishan ZHANG[ab], Jing WANG[a], Zijun ZHUANG[a], Baoping LI[ab*] & Yuan LIU[c]
[a]*Beijing Advanced Innovation Center For Future Education, Beijing Normal University, Beijing*
[b]*Faculty Of Education, Beijing Normal University, Beijing*
[c]*Analytical and Testing Center, Beijing Normal University, Beijing*
*libp@bnu.edu.cn

Abstract: This paper introduces a system that assesses how an elementary student acquires unfamiliar information to solve non-routine real life problems in the perspective of information processing. Students are required to understand the situations, select and apply the relevant information to solve the problems. All the problem solving procedures, including inquiry behaviors, are automatically recorded for off-line data analysis. By running an experiment with a total of 32 grade-5 students, our study evaluated their problem solving processes by analyzing the correlation between students' performance and behavior data. The results reveal that students might have difficulty in applying the information to solve the problems correctly, and it suggests that integrating discrete information may be the biggest obstacle for the elementary students to solve the problems.

Keywords: problem solving, elementary students, information processing, behavior data

1. Introduction

How to solve non-routine problems is one of the most essential skills in 21st century (Griffin, McGaw, & Care 2012). Problem solving is the process of finding a means for reaching some goal from the initial state. Being different from well-educated adults, young students clearly have different levels of competences in solving complex domain-general problems (Findings 2014). It has fostered great demand for teaching this domain-general problem solving skill to them (Greiff et al. 2014).

Fortunately, there has been research focusing on the assessment of problem-solving process. Dickison et al. (2016) assessed nursing clinical judgment by analyzing how problem solvers utilized symptoms of the simulated patient. Schweizer et al. (2013) introduced two instruments MicroDYN and MicroFIN which are proved to be the most well-established tools for assessing complex problem solving ability from the perspective of system thinking. To analyze students' intention based upon behavior data, Evidence-Centered Design (ECD) has to be adopted (Mislevy, 1994), which defines the assessment framework to ensure the way in which evidences are gathered to be able to interpret the underlying assessment purpose. Based on all these previous researches, we aim to provide supports of problem solving for Chinese students who are usually better at mathematics, and worse at integrating discrete pieces of information (PISA, Publishing, 2010; OECD, 2016).

In this study, students are required to first understand the problem, then gather the relevant information to solve it. The key ability of problem solving in our case is to distinguish the information, then apply the appropriate one. On the other hand, as our system does provide more than enough information, the type of information that a student accesses should reflect how the student value the different types of information (Johnson, Häubl, & Keinan 2007).

In this paper, we explained the results that show how elementary students solved problems by applying the relevant material. Our hypothesis is that higher percentage of relevant reading would always lead to better problem solving performance.

2. The Assessment System

The system provides a general framework for assessment tasks and implements the functionalities that support online student interaction recording. In the system, one assessment task could have many test items, and each test item can be either a multiple choice question, fill-in-blank question or an interactive question. Most of the test items cannot be solved without referring to their relevant information, but some standalone test items are existed correlating no relevant document.

To support the structure of the test items, the system provides some common functionalities and utilities. For example, there is a navigation bar placed to the right of the area where the test items display, like Figure 1. All the associated materials to the current assessment task are stored in a component called "material center". Material center contains not only the relevant materials, but also some irrelevant ones. As soon as a student clicks on material center, a new window will popup, which shows the list of all the materials. When a student clicks on the name of a document, the detail will be displayed, like Figure 2. As soon as the student finishes a test item, he/she can proceed to the next one.

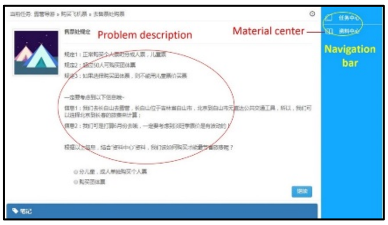

Figure 1. The sample of a test item

Figure 2. material center: The list of the materials (left); The content of a single material (right)

We conducted our experiment and data analysis with one representative assessment task. It uses camping as the story line and contains four test items, as Table 1. In the first test item, students are expected to read the relevant materials to decide which is the best way to purchase tickets. In the second test item, students are expected to correctly measure the length and width of three tents, then use the methods described in the relevant materials to calculate the capacity of the tents. To solve the third problem, students need to assign 54 people into 7 tents, without reading any relevant materials. In the last task, students are required to do a simple calculation for the consumption of water and food for the camping events referring to material center.

Table 1: Summary of test items

Item name/Feature	Item type	Has relevant materials	Type of interactions
Ticket purchase	Multiple choice	Yes	Correct/Incorrect
Tent capacity calculation	Interactive simulation	Yes	Correct/Incorrect
Tent assignment	Interactive simulation	No	Progress/back-step
Path design	Fill-in-blank	Yes	Correct/Incorrect

The system is able to record every single interaction such as clicking on an alternative, and accessing the relevant materials and all the interactions are stored as JSON format into MongoDB. To standardize the interactions, all of them are labeled with 6 different types: correct, incorrect, progress, retreat, valid, and invalid. When the correctness of an interaction can be easily decided given the current context, the interaction will be labeled as correct or incorrect. Otherwise, an interaction will be labeled as progress or back-step based on whether the interaction move the state towards the goal. Some interactive simulations intentionally embed unnecessary interactions which do not help solving the problem at all. In this case, a necessary interaction will be labeled as valid, otherwise will be labeled as invalid.

3. Experiment Design and Analysis Models

The experiment is divided into two parts. First of all, students should go through the introduction task which helps students to become familiar with the User Interface. As soon as the introduction task was done, they started to work on the actual assessment task. Students needed to finish all the test items within 40 minutes independently. 32 elementary students in grade 5 took part in this experiment.

In order to understand how elementary students solve problems in the perspective of information processing and diagnose their issues during the solving procedures, we mined the relationship between information processing behaviors and students' outcome performance in test items. In specific, we inspected their behaviors in terms of their interactions on material center. The behaviors were aggregated as the frequency of relevant and irrelevant reading, and we performed Pearson correlation to measure the relationship. We also conduct Hidden Markov Model (HMM) to infer students' internal problem solving states by observing their behaviors.

The grading criteria is obvious for the first, the second, and the last test item. Students simply earn points when their answers were correct. However, the percentage of the interactions labeled as progress of a student is then used to represent his/her score in the third test item.

When a test item is an interactive simulation, it is possible to get a student's in-progress correctness, which can be also considered as the correctness of a step (Vanlehn et al. 2007). Given this, we extracted this in-progress correctness for the second and the third test items.

4. Results

4.1 Descriptive Results

As mentioned before, there were 4 test items in total. Each item was weighted as 1 point. So the maximum possible score of the test was 4. On average, 36.7% students' reading was relevant to solving the current assessment task. Their average score of the assessment tasks was 1.572 (SD=0.762). The average score and the standard deviation of each individual task was described in Table 2. The table also demonstrated the average frequency of students' relevant reading and irrelevant reading behaviors to finish each item. The results indicated that the difference in how a test item should be answered didn't seem to affect students' performances. However, students performed much better when reading relevant information was not required.

Table 2: Descriptive results of the test items

Task	Ticket purchase	Tent capacity calculation	Tent assignment	Path design
Score	0.5(SD=0.762)	0.25(SD=0.237)	0.68(SD=0.413)	0.136(SD=0.195)
Relevant reading	0.5(SD=0.622)	0.78(SD=0.608)	N/A	0.59(SD=0.911)
Irrelevant reading	0.91(SD=1.254)	0.63(SD=0.793)	0.15(SD=0.330)	1.09(SD=3.306)

4.2 Correlation Results

Pearson correlation coefficients were calculated to discover whether material reading behaviors can affect students' performances. In specific, for each student, we first calculated the overall percentage of relevant material reading behaviors, and calculated Pearson correlation coefficient to check whether this percentage correlates the student's overall performance and the performance in individual test items. The results were showed in Table 3. From the result, different reading behaviors didn't lead to significantly different levels of performances. So we further explore whether their reading behaviors affect in-progress performance. It turned out that correctly referring to the relevant materials could lead to more valid interactions ($r=0.403$, $p=0.022$). Unfortunately, the valid interactions did not always end up with final correctness.

Table 3: Correlation between the percentage of relevant material reading and test item performance

	Ticket purchase	Tent capacity calculation	Tent assignment	Path design	Total score
Pearson correlation	-0.009	0.185	0.305	0.089	0.156
significance	0.962	0.310	0.090	0.627	0.394

4.3 HMM Results

We set the number of hidden states as 3, which represented 3 different competence states. The emission probabilities of the corresponding observations, as listed in Table 4, were used to explain the meaning of the 3 hidden states. A student in H1 was probably bad in solving problems by reading additional materials, because students in this state often read irrelevant materials, and make invalid or incorrect interactions. The observations of H2 were dominated by the behavior “Retreat”, and the observations of H3 were dominated by the behavior “Progress”. The 2 behaviors (“Retreat” and “Progress”) can be only observed in the test item “tent assignment”, which does not require students to read materials. A student in H2 should be bad at solving this planning problem. In contrast, a student in H3 should be good at solving this planning problem. Unfortunately, HMM model failed to detect a hidden state where students stay high competence for solving problems that require additional reading. It is probably because most of the students performed poorly during the assessment.

Table 4: The hidden states and their emission probabilities (drag stay)

Observation /Hidden states	H1 (struggling in information processing tasks)	H2 (struggling in planning)	H3 (smooth in planning)
O1	Invalid & incorrect P=0.178	Retreat P=0.930	Progress P=0.959
O2	Incorrect P=0.163	Progress P=0.0696	Retreat P=0.0274
O3	Valid & correct P=0.152	Invalid, incorrect P<0.001	Long-time reading relevant material P=0.00161
O4	Invalid & correct P=0.123		

The transition probabilities among the hidden states were illustrated in Figure 3. It means that the inferred competence states were relatively stable.

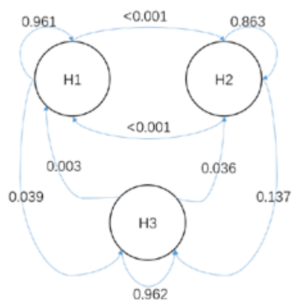

Figure 3. The transition probabilities among the hidden states

5. Discussion and Conclusion

The results showed that students did not perform very well in general. Furthermore, students exhibited low percentage of relevant material reading while performed much better in the task that did not require reading additional materials. It means that integrating discrete information may be the biggest obstacle for the students to solve the problems. The results from the reading section in PISA 2009 (Publishing, 2010) also reported similar findings.

Both of descriptive and HMM results imply that processing outside information to solve problems stands as a relatively independent skill. From the results of HMM analysis showed the students who were good at solving planning problems did not tend to make either relevant or irrelevant readings. In addition, HMM analysis was only able to figure out the hidden state from the data that mainly lead to irrelevant reading and incorrect behaviors. The results again implied that students had difficulty in acquiring and applying relevant information for solving problems. But students did not completely fail in solving the problems. The descriptive result showed that students did go to check the reading materials frequently while necessary, and seldom went to material center when the task did not require, however, at most of the time, students were not able to distinguish which material was relevant. Therefore, students might be able to recognize when they need to refer to additional material, but were at low competence of acquiring relevant information for solving problems.

Acknowledgements

This work was supported by Chinese Ministry of Education Humanities and Social Science Foundation under projects number 14YJC880025, and China Postdoctoral Science Foundation under projects number 2017M610054.

References

Dickison, P., Luo, X., Kim, D., Woo, A., Muntean, W.,... Bergstrom, B. (2016). Assessing Higher-Order Cognitive Constructs by Using an Information-Processing Framework. *Journal of Applied Testing Technology,* 17.

Findings, K. (2014). PISA 2012 results: creative problem solving: students' skills in tackling real-life problems (volume V). *OECD*.

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C.,... Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. [Journal Article]. *Educational Research Review*, 13, 74-83.

Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills:* Springer.

Halverson, R., & Owen, V. E. (2014). Game-based assessment: an integrated model for capturing evidence of learning in play. [Journal Article]. *International Journal of Learning Technology,* 9(2), 111-138.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology Learning Memory & Cognition,* 33(3), 461-474.

Mislevy, R. J. (1994). EVIDENCE AND INFERENCE IN EDUCATIONAL ASSESSMENT, 1 2. *Psychometrika,* 1995(4), 44.

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning & Individual Differences,* 24(2), 42-52.

Vanlehn, K., Koedinger, K. R., Skogsholm, A., Nwaigwe, A., Hausmann, R. G. M., Weinstein, A.,... Billings, B. (2007). What′ s in a Step? Toward General, Abstract Representations of Tutoring System Log Data, 4511, 455-459.

Zhang, L., VanLehn, K., Girard, S., Burleson, W., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y. (2014). Evaluation of a meta-tutor for constructing models of dynamic systems. *Computers & Education,* 75, 196-217.