Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Unpacking Interaction Markers of Critical Thinking

Aditi KOTHIYAL[a+], Rwitajit MAJUMDAR[b*], Shitanshu MISHRA[c], Jayakrishnan Madathil WARRIEM[d] & Prajakt PANDE[e]
[a]*Indian Institute of Technology Gandhinagar, India*
[b]*Kumamoto University, Japan*
[c]*MGIEP UNESCO, India*
[d]*Indian Institute of Technology Madras, India*
[e]*Southern Methodist University, USA*
[+]aditi.kothiyal@iitgn.ac.in, *majumdar@kumamoto-u.ac.jp

Abstract: In this paper, we focus on critical thinking activities conducted in an online environment and describe interaction markers based on the data collected from those activities. The work is based on the ENACT framework. We conducted an empirical study to understand the clusters of critical thinkers based on performance and interactional marker values of the participants (n=37). The paper highlights the definitions of the interaction markers and conducts an clustering analysis of the types of Critical Thinking patterns that emerged. The results show two clusters with similar critical thinking outcome performance but different action patterns. We discuss the need of further empirical evidence relating learning effect on the actions and Critical Thinking.

Keywords: Critical Thinking, Interaction markers, ENaCT framework, embodied cognition

1. Introduction

Critical thinking (CT) skills are widely understood as foundational skills to be available within a person that encompasses cognitive skills, attitudes and dispositions [ref1] and there is a greater focus to integrate them in regular instructional design and practices. There is a wide variety in the approach to inculcate critical thinking skills among practitioners and researchers, but most of them focuses on using authentic or real-world situations to assess skills and dispositions related to problem solving, synthesis and argumentation (ref2). The existing research also covers diverse areas of CT right from defining CT to assessing CT. The work by Scriven and Paul (1987) focuses on conceptualization of CT, by Facione (1994) on CT Skills assessment, by Halpern on a model for teaching and learning of Critical Thinking skills (1998) and the framework by Paul and Elder (2009) are some of the notable examples. These researches do converge in highlighting the need for high-level thinking and the process of analysis, evaluation, reasonableness, and reflection to perform CT activities (Jeevanantham, 2005)

Currently, critical thinking is assessed using standardized tests or rubrics for open ended activities. Some examples include the California Critical Thinking Skills Test (Facione, 1990), the Cornell Critical Thinking Tests (Ennis & Millman, 2005), the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985), and the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980). Researchers have also used rubrics in order to assess students' performance on open ended critical thinking tasks or domain-specific measures of CT (Tiruneh et al, 2014; Mutakinati et al, 2018). These assessments do not consider the thoughts and actions together while commenting about CT (from assessment results) or as Paul and Elder (1998)

We extend the existing work on CT skills and disposition through analysis of interactions between individual elements of thoughts as proposed by Paul and Elder (1998). We collect

interaction logs on a learning environment designed for improving critical thinking and further use them to identify the clusters of critical thinkers based on their interactions.

2. Methodology

The broader goal of this study was to trace CT in terms of thinkers' navigation across, and the interconnections they make between, the different elements of thought presented above. Unlike previous approaches that largely depend on text-based analyses of CT outcomes and processes, this required designing a system (e.g. a digital environment) that explicitly parses the different elements of thought (e.g. as features of the environment), invokes their CT by allowing them to interact with each element (e.g. via its corresponding feature in the environment) as well as (mentally) integrate those elements, and allows examination of one's interaction within that system by helping us capture their interaction behavior which can be eventually mapped on to their CT abilities. The design of such a system was thus integral to our approach.

We adopted a Design-based Research (DBR) framework. DBR emphasizes iterative cycles of design, development, deployment/testing, analysis and redesign of an intervention (Cobb et al., 2003; Wang & Hannafin, 2005), where the theories, design principles, (often technological or technology-enhanced) solutions, and the methods used to evaluate them, systematically evolve across iterations. DBR facilitates a gradual maturation of the research process, products, as well as the involved researchers by leading them to a better understanding of the process of intervention (Amiel & Reeves, 2008).

2.1 ENACT environment

To seek answers to this question, we created a web-based critical thinking environment with information and expression affordances to take a critical decision in a given scenario.

Task description: This is a panel at the top of the screen which describes the situation in which the solver must think critically and make a decision. This task gives the solver a context for critical thinking.

Information affordances: On the left side of the screen are a set of panels which contain information (data and concepts) related to the task. Data is presented in multiple formats such as bar graphs, line graphs and coverage maps, and the solver is free to use them as they wish to do the task, with the constraint that only one panel opens at a time.

Expression affordances: On the right side of the screen, are a set of panels with textboxes, labelled purpose, question, assumptions, conclusions and implications as per the "Elements of Thought" framework. Again, the solver may input text whenever they wish, with the constraint that only one textbox opens at a time. The solvers input their final decision in the implication's textbox.

Summary panel: At the bottom of the screen, is a panel where all the information the solver entered in the five textboxes is collated and presented together.

Users are given a specific critical thinking task which involves responding to a focus issue while integrating information from multiple representations. The responses were scaffolded according to the Paul and Elders (2008) elements of thoughts as prompts in the expression affordances. The environment enabled logging of all the click interactions when the learner accessed any of the information affordances on the left or the expression affordances on the right. At a given instance, only one information affordance and one expression affordance are seen in the accordion of the menus. When the user enters the task page, the interaction log is recorded and sets it as the time to start the task. Once the user presses the done button, the task is considered to be complete.

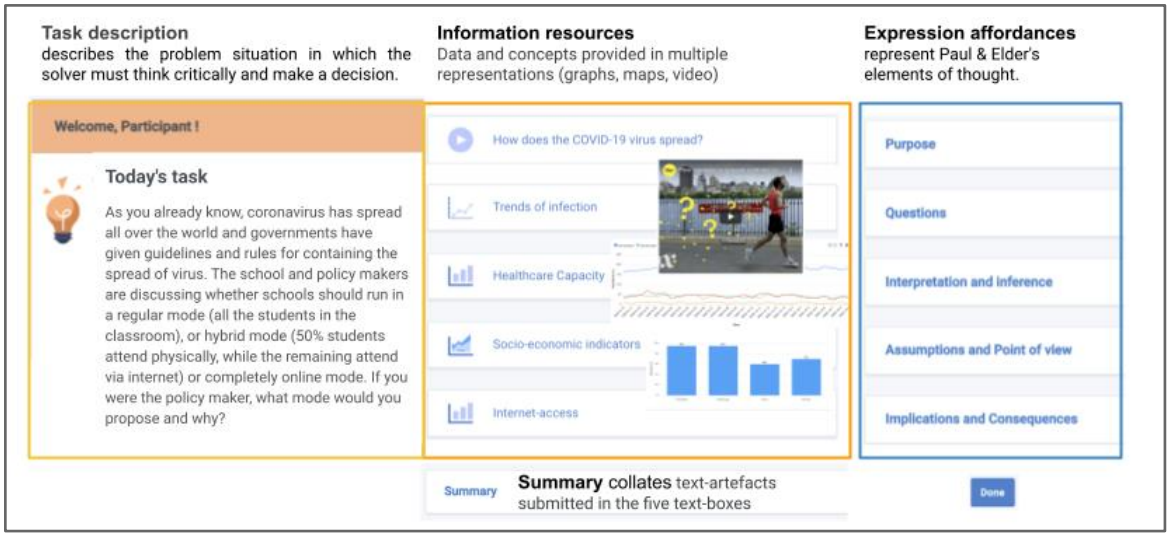

Figure 1. Critical thinking task environment (Majumdar et al. 2021)

3. Research Study

3.1 Research Question

In this work our research question is: *What is the relationship between performance on a critical thinking test and one's navigation across and interconnections between the different elements of thought of critical thinking?*

3.2 The study

We conducted an online study and invited participants aged 10 years and up with various educational backgrounds. Participants consented to logging their data within the ENaCT system and received a gift card of INR 200 for their participation. The study was set-up within the Moodle environment and consisted of three parts: a pre-test, a decision-making problem within the ENaCT environment and a post-test, which consisted of the same questions as the pre-test but in a different order. The participants were given login credentials and had a maximum of 1 hour to complete the study at a time of their convenience. For young children under the age of 15, we set up synchronous Zoom sessions during which participants could do the study and a researcher was available to provide technical support if necessary. Since the experiment was largely unmonitored, participants sometimes missed doing one of the tests or spent longer than the allotted time on the activity. We removed such participants from our data set and ultimately, we had data from 37 participants (21 females; aged: M=12.2 years, SD 1.3 years) at school level.

3.3 Assessment instrument

To measure critical thinking, we designed an assessment instrument (see Appendix A) which consists of seven items, one for assessing each element of thought namely, purpose, concept, information, conclusions, assumptions, perspectives and implications. The test was based on the Watson Glaser critical thinking test (Watson & Glaser,1980). Each item consisted of a scenario, followed by a few questions based on the scenario assessing the respective element of thought. The test was piloted with 15 participants to ensure the understandability of the items and finalize the grading rubric. In this study, the responses of all participants were graded by two researchers and any differences were resolved through discussion to ensure complete agreement in the grading.

3.4 Analysis approach

Our analysis approach follows directly from our research question which focuses on exploring the relationship between navigation actions and critical thinking performance. Therefore, we define metrics of interaction with the system that we hypothesize on the basis of critical thinking literature to have an impact on the decision-making activity in ENaCT. The actions that an individual can do within the ENaCT system include looking at the information panels (called *info* actions), viewing the element of thought questions (called *expr* actions), entering responses to the elements of thought questions or their final decision (called *sub* actions) and viewing a summary of their entered responses and decision (called *chksum*). However, the solver can do these actions any number of times and in any sequence. So, we differentiate between an action being done for the first time versus being repeated because the role of each of these actions in decision making would be different.

We conjecture that there exists a relationship between solver's actions within the ENaCT system and their critical thinking performance as measured by the test, ie, some interactions and levels of performance "go together". Further, we conjecture that there exist different profiles of critical thinkers that differ in terms of their actions and critical thinking performance. Concretely what this means is that different kinds of critical thinkers engage with the elements of thought to different extents and in different sequences, and this results in different levels of critical thinking performance. In order to identify these different profiles, we consider actions and performance metrics together in the analysis pipeline shown below.

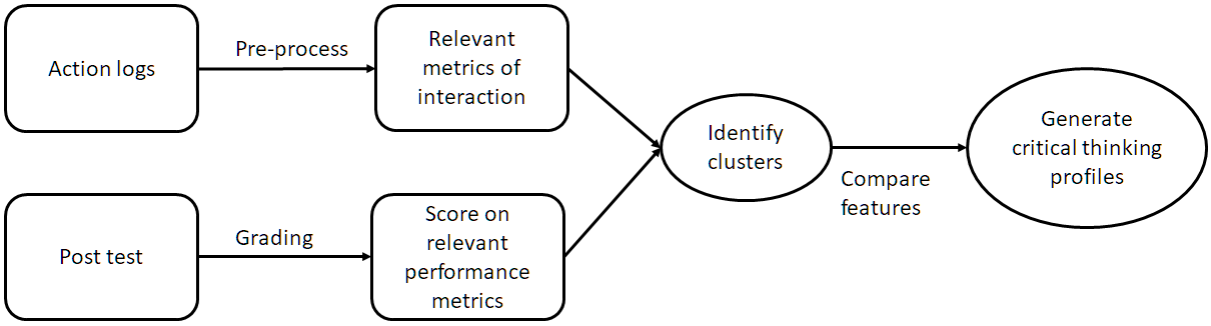

Figure 2. The relationship between actions and critical thinking performance at the data level

3.4.1 Metrics of interaction and performance

The first step in our analysis is to define theoretically relevant metrics of interaction as seen below:

1. ***Ratio_info_actions/Ratio_expr_actions/Ratio_sub_actions/Ratio_chksum_actions****:* These are the ratios between the number of info/expr/sub/chksum actions and the total number of actions performed. These metrics are relevant because they provide quantitative measures of different cognitive processes known to be productive in decision-making such as,
    a. Information gathering *(info)* from multiple external sources and representations of information which interact with internal representations of information (Kirsh, 2010, Hutchins et al 2013),
    b. Viewing the elements of thought questions or reading previously written responses to these questions *(expr)* as a way of cueing or stimulating further thoughts (Kirsh, 2009, 2010),
    c. Entering responses to the elements of thought questions or their final decision *(sub)* which helps to create persistent referrents that can facilitate further inquisition and be re-represented (Kirsh, 2009, 2010; Hutchins, 2005)
    d. Viewing their previously written thoughts *(chksum)* as a way of monitoring, evaluating and regulating their work (Kirsh, 2004).

2. ***Ratio_revinfo_actions/Ratio_revexpr_actions/Ratio_resub_actions:***
   These metrics are the ratios between the number of repetitions of any *info/expr/sub* actions and the total number of actions. We consider these relevant interaction metrics because revisiting an information panel or an expression panel or resubmitting is an indication of reflective behaviours such as monitoring and evaluation, which is known to be productive in problem-solving (Kirsh, 2004).
3. ***nInter_expr_info/nInter_info_expr/nInter_sub_info/nInter_info_sub:***
   These metrics are the number of times an individual goes from an information panel to an expression panel and vice-versa, or an information panel to a submission panel and vice-versa. We consider the interleaving of these actions because information gathering and reading or writing about the elements of thought are the two main affordances of our interface and the interleaving of these actions is an indicator of an attempt to "integrate" the two affordances during decision-making.
4. ***Conclusion/Implications/Assumptions score***:
   The scores obtained by individuals on the conclusions, implications and assumptions elements of thought of the post-test are considered as indicative of their critical thinking performance. We only consider the scores on these three elements of thought because *Conclusions* and *Implications* represent the result of the students' reasoning and *Assumptions* are the basis for this result. Together these three elements evaluate the result of the students critical thinking reasoning.

Together we have fourteen features representing the critical thinking of our participants. Different values on metrics of type 1, 2 and 3 will represent different navigation patterns and we conjecture that these will differentiate good and weak critical thinkers as identified by their score on the post-test. Specifically, we hypothesize that:

1. Lower values of type 2 ratios (*Ratio_revinfo_actions/ Ratio_revexpr_actions/ Ratio_resub_actions)* will constitute poor critical thinking performance as they indicate lower evaluation and revision, while higher values of type 2 ratios will constitute good critical thinking performance.

2. Within good critical thinkers we conjecture that there would be multiple approaches to critical thinking, one which interacts with the provided context-related information more and one which interacts with the cues of the elements of thought more. However, we conjecture that good critical thinkers would have a higher proportion of interleaving actions of the type *info-expr* or *expr-info* as these represent an attempt at integration.

With these hypotheses we perform clustering on the fourteen features using the k-means algorithm. Before clustering, the features are normalized using a min-max scaler to lie between 0 and 1 because the range of the various features is different. The number of clusters is chosen to simultaneously minimize the inertia and maximize the silhouette of the chosen number of clusters. In order to examine the difference between the clusters in terms of the critical thinking features, we performed pair-wise Kruskal-Wallis tests. This enables us to identify the significantly differing features between the clusters, and thus build meaningful and interpretable profiles of critical thinkers based on their features.

4. Results

4.1 Performance on each element of thought

We first report the post-test scores of participants on all elements of thought in Figures 1-7 below in order to understand the performance of the participants. As we see from the scores, the participants scored highest on the elements of *Concept* and *Perspective*. The lowest scores are for the elements of *Conclusions, Assumptions* and *Implications,* suggesting that participants found it difficult to make decisions with reasonable assumptions and desirable consequences.

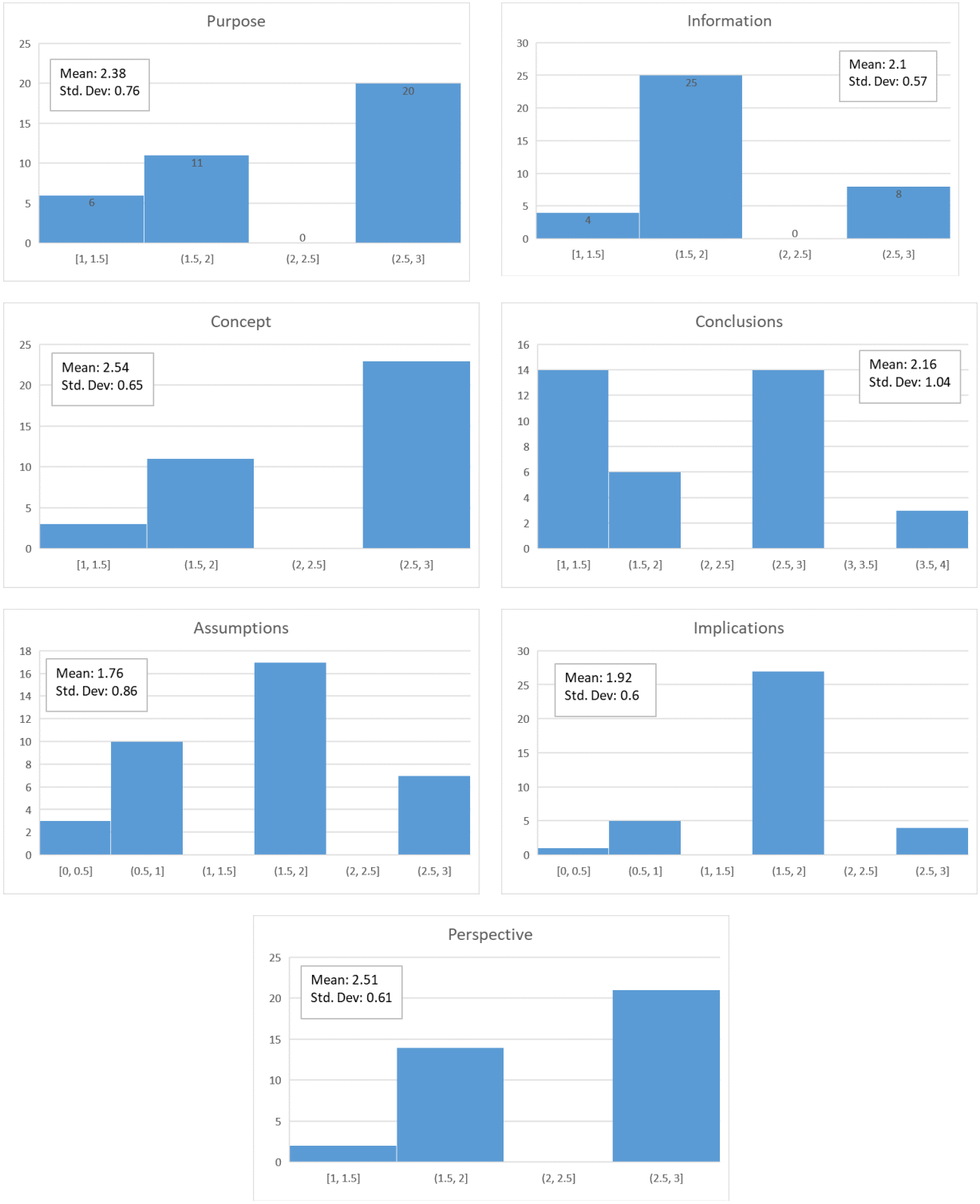

Figure 3. Distribution of the Performance score in each element of thought

4.2 Interaction with each element of thought

Next, we report the descriptive statistics of the 14 interaction features that we consider for generating critical thinking profiles and defined in Section 4.4.1. We find that on average participants do most information gathering actions, followed by actions of reading the elements of thought cues. The number of chksum actions is the lowest, which means participants did not check their responses, indicating poor reflective behaviors. Among the revisits, we again find that the average is the highest for revisits

to the information panels, followed by the expression panels. Finally, the number of interleaving actions, where participants did an *info* action followed by an *expr/sub* action, or vice-versa, was close to one, suggesting that participants did not interleave much, shifting from one “side” to the other “side” only once while completing the problem. This indicates an overall pattern of the type looking at the information, moving to the expression panels, moving back to the information, making submissions and sometimes returning to the information panels. Such a pattern of interaction reiterates the lack of reflection suggested by the low *chksum* actions.

4.3 Characteristics of profiles of critical thinking

In order to identify profiles of critical thinking performance, we performed cluster analysis using the 14 features of participants' interaction in the ENaCT system defined in Section 3.4.1. We identified four clusters and the average values of the features in each of the clusters is shown in Figure 4. The p-values of Kruskal-Wallis tests comparing the statistical difference between the features of each pair of clusters shows that several of the interaction features of each pair of clusters are statistically different, indicating that the clusters represent distinct critical thinking behavior profiles. Specifically, we see that cluster two has a significantly lower conclusion score than all other clusters, while cluster two also has a significantly higher implications score than cluster zero. In the interaction features we see that cluster one has significantly higher information actions and information revisits than clusters zero and two. Similarly, cluster one also has significantly lower expression actions compared to all other clusters, together suggesting that cluster one is an information-favoring cluster. Cluster two has significantly higher *chksum* actions compared to all other clusters suggesting that this is a more reflective cluster. In terms of interleaving actions, cluster three has the significantly higher information-expression interleaving than the other clusters and cluster one has significantly higher information-submission interleaving than other clusters, both of which indicate integrative interface exploration. These differences between the clusters are further depicted in Figure 4 and each of the profiles built on the basis of these differences is elaborated below.

1) *Cluster 0 - Non-exploring, instinctively good deciders:* Cluster 0, which has 11 participants, is characterized by high Conclusions and Assumptions scores and medium scores on Implications. In terms of interaction, participants in this group do few information actions, but a high number of expression, submission and resubmission actions. Apart from this they had low values for every other interaction metric including revisits and interleaving. This suggests that these participants were good, but *instinctive deciders*, preferring to rely more on their prior knowledge and experiences to make a decision rather than exploring and integrating the resources in the interface. Further these participants were also non-reflective in that they did not look at their submitted responses, possibly because they were confident in their decisions.
2) *Cluster 1 - Information-exploring moderately good critical thinkers:* Cluster 1 has nine participants who obtained medium scores on Conclusions, Assumptions and Implications. These participants had high values of information actions, including revisits, but low values on expression, submission and resubmission actions. Further these participants had medium to high values of *chksum,* revisits to expression panels and all interleaving actions, specifically information-submission interleaving. Together these findings suggest that these participants were moderately good critical thinkers, who explored and used the information on the interface to write their responses to the elements of thought. Even though they

showed some reflective behaviors by looking at the summary of their submitted responses and re-looking at the expression panels, they relied primarily on the given information and so we label them *information-exploring*.

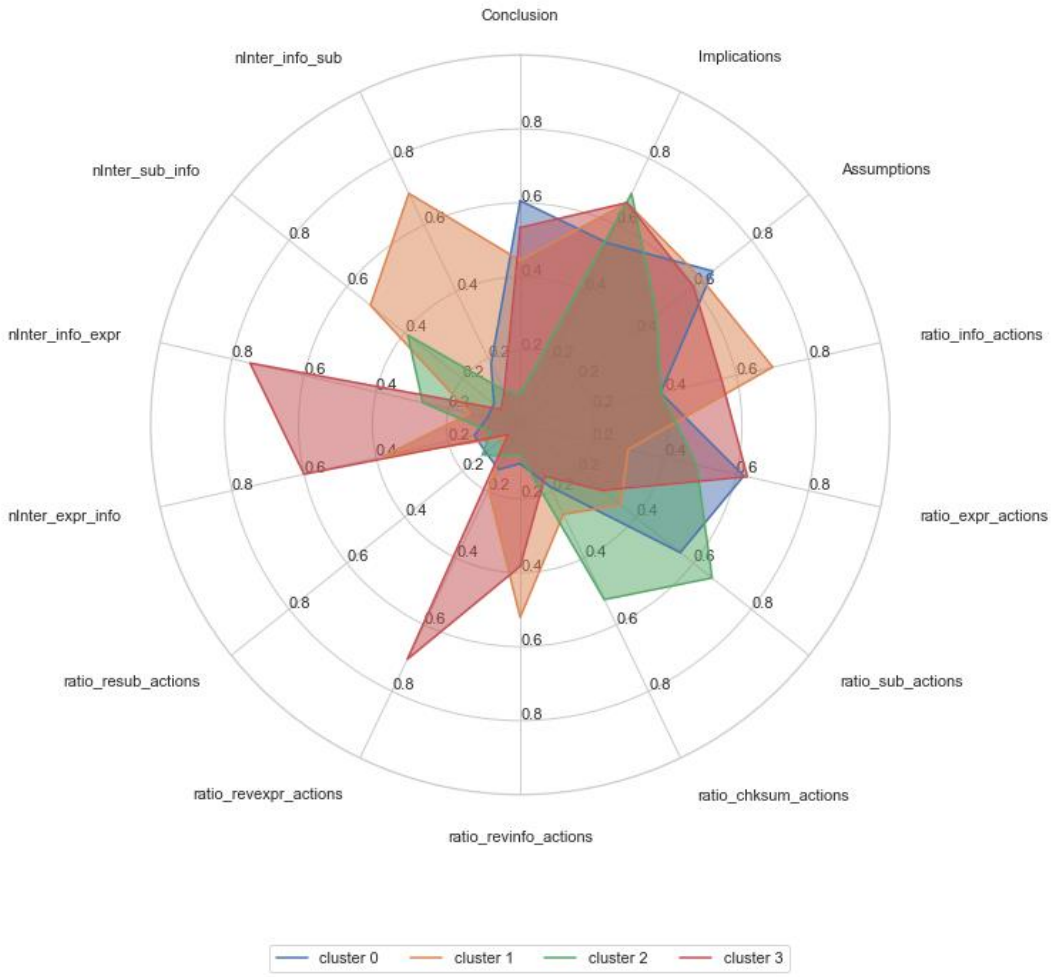

Figure 4. Clusters of critical thinking behaviors

3) *Cluster 2 - Reflectively-exploring poor deciders:* This cluster is the largest with 12 participants, who have the lowest score on Conclusions and Assumptions, but the highest in Implications. Among interaction features, these participants have low-medium values for information and expression actions, including revisits, but the highest value for submission and resubmission actions. Further they have the highest value for *chksum* actions, and low to medium values for interleaving actions. Taken together, it appears as if this group were not very good at making a decision, but good at understanding its consequences. However, their behavior is highly reflective as seen in their high number of resubmissions and looking at the summary of their responses frequently. Finally, they did explore and integrate the interface features fully, because of which we label them *reflectively exploring.*
4) *Cluster 3 - Comprehensively-exploring moderately good critical thinkers*: This is the smallest cluster with only five participants who scored medium on Conclusion, Assumptions and Implications. In terms of interaction, they had medium-high values for both information and expression actions and revisits, but the lowest values among all clusters for submissions, resubmissions and *chksum* actions. However, they had the highest values for expression - information interleaving and information - expression, and the lowest values for information - submission and

submission - information interleaving. Together these findings suggest a pattern of exploring and integrating the interface features comprehensively, but this exploration did not lead to many responses and was not reflective, along with moderately good critical thinking performance. For these reasons, we label this cluster *comprehensively-exploring*.

5. Conclusion

This study focused on defining the interaction indicators of critical thinking based on elements of thoughts defined by Paul and Elder (2009) and the ENaCT framework (Mishra et al. 2020). An empirical study was conducted with participants from the school level (10 years and above) to execute a critical thinking task (Majumdar et al. 2021) and collected their interactions and critical thinking outcomes. Extending the previous analysis of sequential pattern mining from the interactions to identify CT Processes (Mishra et al. 2021), this study did an analysis of comparing interaction indicators with the CT outcomes. The analysis showed that there can be clusters with similar CT outcome but different action patterns and results demonstrate different kinds of relationships between action and CT - one with evidences of some learning effect. Neither is this entirely unknown in education and/or problem-solving research. The learning effect was unexpected but perhaps not entirely surprising as learning is inevitable irrespective of the problem solvers' expertise and the complexity of the problem/context. The learning effect is good as it has implications to designing for learning CT.

Finally, the influence of learning effect on the relationship between action and CT needs a separate detailed investigation. This is also where the 4E cognition and learning paradigms may be most useful.

Acknowledgements

The research work was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) JP22H03902 and international collaboration grant SPIRITS 2020 of Kyoto University. The authors extend their sincere thanks to the participants of the study and Ms. Kinnari Gatare for supporting data collection and Dr. Huiyong Li for the ENaCT system configuration.

References

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of educational technology & society*, *11*(4), 29-40.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, *32*(1), 9-13.

Ennis, R. H., & Weir, E. E. (1985). The Ennis-Weir critical thinking essay test: An instrument for teaching andtesting. Midwest Publications.

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). Cornell critical thinking tests: Administration manual.California: Critical Thinking Company.

Facione, P. A. (1990). The California Critical Thinking Skills Test--College Level. Technical Report# 1. Experimental Validation and Content Validity. https://files.eric.ed.gov/fulltext/ED327549.pdf

Facione, P. A. (1994). Using the Holistic Critical Thinking Scoring Rubric to train the discovery of evidence of critical thinking. *Critical Thinking and Clinical Reasoning in the Health Sciences*.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American psychologist*, *53*(4), 449.

Halpern, D. F. (2010). Manual HCTA, Halpern Critical Thinking Assessment (Version 2.1). Moedling, Austria:Schuhfried

Jeevanantham, L. S. (2005). Why teach critical thinking? *Africa Education Review*, *2*(1), 118–129. https://doi.org/10.1080/18146620508566295

Kirsh, D. (2010). Thinking with external representations. *AI & society*, *25*, 441-454.

Majumdar, R., Kothiyal, A., Mishra, S., Pande, P., Li, H., Yang, Y. Y., & Warriem, J. M. (2021, July). Design of a critical thinking task environment based on ENaCT framework. In *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp. 294-296). IEEE.

Mishra S., Majumdar R., Kothiyal A., Pande P., Warriem J.M. (2020) ENaCT: An Action-based Framework for the Learning and Analytics of Critical Thinking. in the LA@ICCE workshop of ICCE 2020

Mishra S., Majumdar R., Kothiyal A., Pande P., Warriem J.M. (2021) Tracing Embodied Narratives of Critical Thinking. In: Roll I., McNamara D., Sosnovsky S., Luckin R., Dimitrova V. (eds) *Artificial Intelligence in Education.* AIED 2021. Lecture Notes in Computer Science, vol 12749. Springer, Cham. https://doi.org/10.1007/978-3-030-78270-2_48

Paul, R., & Elder, L. (2009). Critical thinking, creativity, ethical reasoning: A unity of opposites. *Morality, ethics, and gifted minds*, 117-131.

Scriven, M., & Paul, R. (1987, March). Critical thinking. In *The 8th Annual International Conference on Critical Thinking and Education Reform, CA* (Vol. 7, No. 9).

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, *4*(1), 1-17.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, *53*(4), 5-23.

Watson, G. (1980). Watson-Glaser critical thinking appraisal. San Antonio, TX: Psychological Corporation.