Validity of Concept Mapping for Assessing Mental Models of System Functioning

Judith Schmidt (judith.schmidt@tu-dresden.de)[1,2]

Eva Louise Rix (eva_louise.rix@tu-dresden.de)[1]

Peter Hesse (peter.hesse@tu-dresden.de)[1]

Stephan Abele (stephan.abele@tu-dresden.de)[1]

Romy Müller (romy.mueller@tu-dresden.de)[2]

[1]Chair of Vocational Education, TUD Dresden University of Technology
[2]Chair of Engineering Psychology and Applied Cognitive Research, TUD Dresden University of Technology

Abstract

Having a correct mental model of a technical system facilitates interaction and problem solving. To assess such mental models of system functioning, appropriate methods are needed. We tested whether concept mapping with a focus on means-ends relations leads to valid assessments of participants' mental models of system functioning. Automotive and utility vehicle apprentices constructed concept maps of two simple, everyday systems (bike, traffic) and one complex, domain-specific system (fuel temperature control). However, only one group of participants had previously covered the complex system in class. Aspects of participants' concept maps regarding content (correct functional propositions) and structure (intersection over union) were assessed and related to respective reference maps. Results indicated that group differences in knowledge about the complex system were represented by concept map content, but not structure. We argue that the applied structural reference might need to be adapted to match typical requirements of the domain and task.

Keywords: mental model assessment; concept mapping; system functioning; functional abstraction

Introduction and Theoretical Background

Mental models reflect an individual's understanding of a particular aspect of the world, making it possible to anticipate events and decide how to respond to them (Johnson-Laird, 2013; Rouse & Morris, 1986). Consequently, mental model quality is closely related to performance (e.g., Bußwolder, 2015; Gary & Wood, 2011; Mumford et al., 2012; Sarter, Mumaw, & Wickens, 2007), and mental models differ between groups with varying knowledge and proficiency regarding a certain topic (e.g., Chi, Feltovich, & Glaser, 1981; Hmelo-Silver, Marathe, & Liu, 2007; Jee, Uttal, Spiegel, & Diamond, 2015). Thus, mental models are thought to be an important basis for successfully responding to relevant requirements. To examine relations between performance and mental models, these models first must be assessed. While mental model quality is often assessed with proxies such as questions about system functioning, these measures do not provide insights into mental model structure.

One method that represents mental model contents as well as the relations between them is concept mapping. Concept maps are graphical representations in which concepts are connected via directed and labeled links. These link descriptions specify the type of association between the concepts (e.g., functional, physical, or time-related). Individual concept-relation-concept units (so-called propositions) represent the smallest meaningful structures within a concept map and often form the basis for assessment. To assess differences in understanding, concept maps constructed with a pre-determined set of concepts but free choice of linking phrases are most appropriate (Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Ruiz-Primo, Shavelson, Li, & Schultz, 2001).

Of course, a reference for assessing concept map content and structure is also required. Several frameworks for representing system functioning are available, one of which being Rasmussen's (1986) abstraction hierarchy. This hierarchy describes system functioning on several levels of abstraction. The purpose is specified by the highest level, functions by middle levels and concrete physical implementations by lower levels. Since adjacent levels are connected via means-ends relations, the next highest level describes why a certain function or component is implemented, while the next lowest level describes how this is achieved. The abstraction hierarchy has successfully been applied to different systems, such as flexible manufacturing (Bennett, Edman, Cravens, & Jackson, 2023), healthcare (St-Maurice & Burns, 2017), chemical processes (Ade, Peres, Sasangohar, & Son, 2019; Bisantz & Vicente, 1994), and military control (Burns, Bisantz, & Roth, 2004). It can also be used to conceptualize and analyze behavior in these systems. For example, successful problem-solving is characterized by using information from different levels of abstraction (Burns, 2000; Feuerstack & Saager, 2022), whereas unsuccessful problem-solving is characterized by a fixation on particular levels (Hall, Rudolph, & Cao, 2006). Information about the means-ends relations seems to be especially relevant for interacting with complex systems (Borst, Bijsterbosch, van Paassen, & Mulder, 2017; Burns, 1999). In sum, the abstraction hierarchy is relevant for system interaction as it can be used to map system functioning.

Combining concept mapping with the abstraction hierarchy might be a promising method for assessing mental models of system functioning. Participants could be provided with concepts on different levels of abstraction and be instructed to specify the means-ends relations between them. However, can the resulting representations be used to assess mental models of rather complex, domain-specific systems? For example, participants might be able to construct a good concept map by simply applying the rules of the abstraction hierarchy (i.e., connect concepts on adjacent levels of abstraction using means-ends relations), even though they do not know about system functioning.

To address the above question, apprentices from a technical domain (automotive and utility vehicle mechatronics) constructed concept maps about simple, everyday systems (bike, traffic) and a more complex, technical system (fuel temperature control). Two groups with varying knowledge about system functioning were compared regarding the content and functional structure of their concept maps. If the concept mapping method leads to valid mental model assessments, concept maps about the simple systems should be of high-quality content and structure, and there should be no differences between the groups. For the complex system, concept maps constructed by participants knowing the system should be of higher quality.

Method

Participants

Eighty-one automotive and utility vehicle apprentices in their second or third year of training were recruited from three vocational schools in Germany. Since the concept mapping task was part of a larger study protocol, 20 participants were excluded because their complex concept map was not about fuel temperature control. Another 3 participants were excluded because they did not participate in the concept mapping task. The final sample consisted of 58 automotive and utility vehicle apprentices (19 in their second and 39 in their third year of training) who constructed at least one concept map about a simple system and the concept map about the complex system. The demographic information is reported in Table 1. Due to technical difficulties during data collection in school 1, only an age range can be reported. For schools 2 and 3, one participant did not provide demographic information, respectively. All participants gave informed consent.

Table 1: Demographic information.

	N	Year of training	Male/ female	*M* age (*SD*)	Range age
School 1	19	2nd	-	-	18-31
School 2	15	3rd	14/0	19.7 (1.3)	18-22
School 3	24	3rd	22/1	21.0 (3.5)	19-35

Apparatus and Stimuli

Technical Setup The study was conducted directly in the vocational schools during classes. Participants were provided with laptops for viewing the instruction videos and for constructing the concept maps with CmapTools (Florida Institute for Human & Machine Cognition, 2019).

Instruction Participants watched three instruction videos (4-6 minutes each) that explained the abstraction hierarchy, concept map construction, and means-ends relations. Participants were not allowed to add concepts, but could leave out concepts if they did not know how they contributed to system functioning. They were encouraged to begin their concept map with the goal and then proceed along the abstraction levels, and to use arrows to indicate the direction in which the proposition should be read. Finally, participants were asked to describe the functional relations in their own words.

Reference Maps and Provided Materials For each topic (i.e., bike, traffic, fuel temperature control), a reference map containing only correct functional relations was developed to evaluate participants' concept maps (see Figure 1 for the bike reference map). The structure of the reference map followed the abstraction hierarchy. Participants received all concepts contained in the reference maps via CmapTools (see Table 2 for the number of concepts and propositions). All concepts were colored to indicate their level of abstraction (i.e., goal, function, component, or property). As is evident from Table 2, slightly different versions of the reference map for fuel temperature control were used. Since data collection started with the second-year apprentices, an earlier version of the reference map was used. In the later version, two concepts were added to decrease the number of propositions required to fully explain system functioning.

Table 2: Reference map characteristics.

Reference map	Concepts	Propositions
Bike	13	14
Traffic	13	15
Fuel temperature control 2nd year	20	31
Fuel temperature control 3rd year	22	27

Procedure

Participants first watched all instruction videos and were allowed to ask questions. After the first video, a class exercise was included in which participants volunteered to read an example concept map out loud. Following the instruction, participants constructed the first simple map, explaining the functioning of a bike. After 20 minutes, they constructed the second simple map, explaining the functioning of a traffic system. After 15 minutes, they constructed the complex map, explaining the functioning of fuel temperature control. Since participants received more material for constructing the final map, 30 minutes were allotted for this task. The entire procedure lasted one 90-minute class.

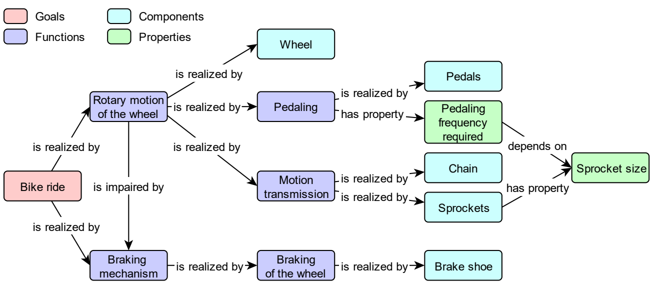

Figure 1: Bike reference map.

Data Analysis

Propositions The number of propositions (i.e., concept-relation-concept units) was used to indicate the size of the concept map. Since the number of provided concepts differed between concept maps, a standardized value was calculated to make these maps comparable. This was achieved by dividing the number of propositions in the participant's map by the number of propositions in the reference map.

Proposition Types To characterize the propositions used to explain system functioning, each proposition was coded as functional (e.g., motion transmission – is realized by – chain), non-functional (e.g., pedals – are connected to – wheel), or unspecified (no relation description given). The coding scheme was developed by the first, second, and last author prior to data analysis. The coding itself was carried out by the first author, and ambiguous cases were discussed between the authors to reach a consensus. For each concept map, the proportion of each proposition type was determined.

Correct Functional Propositions To indicate the quality of concept map content, the number of correct functional propositions was determined. Functional propositions were coded as correct when the presented content could be interpreted as correct, even if it was not represented in the reference map. For example, the reference bike map states that the rotary motion of the wheel is realized by pedaling and motion transmission, which are in turn realized by pedals, chain, and sprockets. A participant may have skipped the second functional level and stated that the rotary motion of the wheel is realized by the chain. This would be coded as correct, even though the functional relation between the rotary motion and the chain is mediated by other concepts in the reference map. Consequently, the number of correct functional propositions does not represent an overlap between the participant's and the reference map, but merely the number of propositions that would be considered to represent existing functional relations. The coding was carried out by the first author, and ambiguous cases were discussed between the authors to reach consensus. Similar to the number of propositions, a standardized value was calculated by dividing the number of correct functional propositions in the participant's map by the number of propositions in the reference map.

Originally, the direction of the arrow connecting two concepts was not included in determining the accuracy of a functional proposition. To analyze whether the decision to omit the arrow direction changed the results, a second accuracy coding was applied that did consider this direction.

Intersection over Union The quality of mental model structure was indicated by the intersection over union (cf. Goldsmith & Davenport, 1990) representing the alignment between a participant's map and the reference map. A maximum value of 1 indicates that the connections in the participant's map and the reference map are identical. The intersection over union decreases when participants omitted connections from the reference map, added new connections, or both. A minimum value of 0 indicates that participants included none of the connections from the reference map.

Statistical Analyses For all dependent variables, a mean value for the two simple systems was calculated. Two participants only constructed a concept map for one simple system and therefore, these values were used directly. We statistically analyzed propositions (standardized), correct functional propositions (standardized), and intersection over union using 2 (*complexity: low, high*) x 2 (*year of training: second, third)* mixed-design ANOVAs with the within-subjects factor complexity. Further, we statistically analyzed the proportions of proposition types using a 2 (*complexity: low, high*) x 2 (*year of training: second, third*) x 3 (*proposition type: functional, non-functional, unspecified*) mixed-design ANOVA with the within-subjects factors complexity and proposition type. If the sphericity assumption was violated, a Greenhouse-Geisser correction was applied and the degrees of freedom were adjusted. All pairwise comparisons used Bonferroni correction.

Results

Propositions

On average, maps contained an absolute number of 11.7 propositions for the simple systems and 18.6 propositions for the complex system (see Figure 2). A standardized value was calculated to be used in the statistical analysis (see section Data Analysis for details). This analysis revealed a significant main effect of complexity, $F(1, 56) = 43.0$, $p < .001$, $\eta^2 = .43$. Thus, participants used more propositions for explaining the simple compared to the complex system (.81 vs. .66, respectively). There was no significant main effect of year of training, $F(1, 56) = .7$, $p = .39$, $\eta^2 = .01$, but a significant interaction between complexity and year of training, $F(1, 56) = 25.9$, $p < .001$, $\eta^2 = .32$. Participants in their second year used more propositions for explaining the simple compared to the complex system (.88 vs. .53, respectively), $p < .001$. For participants in their third year, propositions did not differ between low and high complexity (.77 vs. .73, respectively), $p = .21$ (see Table 3 for mean values).

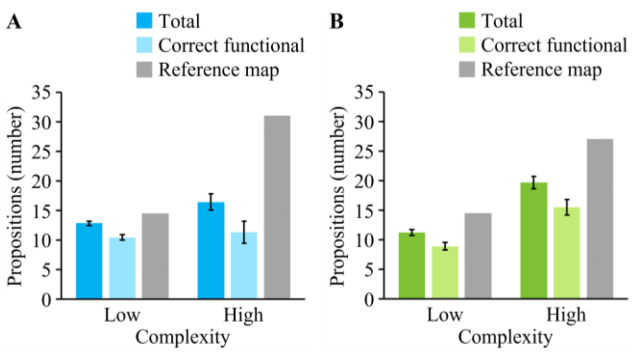

Figure 2: Propositions and correct functional propositions for participants in their (A) second and (B) third year of training.

Table 3: Descriptive statistics for dependent variables.

		Low complexity		High complexity	
Dependent variable	Year	*M*	*SD*	*M*	*SD*
Propositions (number)	2nd	12.8	1.6	16.4	5.8
	3rd	11.2	3.3	19.7	6.4
Propositions (standardized)	2nd	.88	.11	.53	.19
	3rd	.77	.25	.73	.24
Functional propositions (%)	2nd	92.9	9.9	72.7	40.8
	3rd	85.0	28.4	84.4	32.3
Non-functional propositions (%)	2nd	3.7	7.4	3.4	7.4
	3rd	0.9	3.2	1.0	3.3
Unspecified propositions (%)	2nd	3.4	8.2	23.9	41.3
	3rd	14.1	28.6	14.6	32.5
Correct functional propositions (number)	2nd	10.4	2.0	11.3	8.0
	3rd	8.9	3.9	15.4	8.2
Correct functional propositions (standardized)	2nd	.72	.14	.36	.26
	3rd	.61	.27	.57	.30
Intersection over union	2nd	.43	.13	.16	.08
	3rd	.40	.15	.14	.07

Proposition Types

Since the proportions of the different proposition types sum up to 100 for both levels of complexity, the main effect of complexity is not interpretable and not reported. The analysis revealed a significant main effect of proposition type, $F(1.04, 58.38) = 132.66$, $p < .001$, $\eta^2 = .70$. Concept maps generally contained a higher proportion of functional than non-functional or unspecified propositions, all *p*s < .001, and a higher proportion of unspecified than non-functional propositions, $p = .01$. There was no significant main effect of year of training, $F(1, 56) = 1.53$, $p = .22$, $\eta^2 = .03$.

The analysis further revealed a significant two-way interaction between complexity and proposition type, $F(1.03, 57.90) = 7.79$, $p = .007$, $\eta^2 = .12$. Concept maps explaining the simple systems contained a higher proportion of functional propositions, $p = .007$, and a lower proportion of unspecified propositions, $p = .006$. The proportion of non-functional propositions did not differ between complexity levels, $p = .81$. There was no significant interaction between complexity and year of training, $F(1, 56) = .16$, $p = .69$, $\eta^2 = .00$, or the year of training and proposition type, $F(1.04, 58.38) = .10$, $p = .91$, $\eta^2 = .00$.

Finally, there was a significant three-way interaction between complexity, proposition type, and year of training, $F(1.03, 57.90) = 6.98$, $p = .010$, $\eta^2 = .11$. Participants in their second year included a higher proportion of functional and a lower proportion of unspecified propositions to explain the simple compared to the complex system, both *p*s < .002. The proportion of non-functional propositions did not differ between complexity levels, $p = .71$. For participants in their third year, none of the proposition types differed between complexity levels, all *p*s > .89 (see Table 3 for mean values).

In sum, functional propositions were generally most prevalent, whereas non-functional propositions occurred least frequently. Participants who were less familiar with the complex system included fewer functional and more unspecified propositions in this concept map.

Correct Functional Propositions

On average, maps contained an absolute number of 9.4 correct functional propositions for the simple systems and 14.1 for the complex system (see Figure 2). A standardized value was calculated to be used in the statistical analysis (see section Data Analysis for details). This analysis revealed a significant main effect of complexity, $F(1, 56) = 38.8$, $p < .001$, $\eta^2 = .41$. Generally, correct functional propositions were more frequent in concept maps explaining the simple compared to the complex system (.65 vs. .50, respectively). There was no significant main effect for year of training, $F(1, 56) = .6$, $p = .45$, $\eta^2 = .01$, but a significant interaction between year of training and complexity, $F(1, 56) = 24.0$, $p < .001$, $\eta^2 = .30$. For participants in their second year, correct functional propositions (standardized) were more frequent in concept maps explaining the simple compared to the complex system (.72 vs. .36, respectively), $p < .001$. For participants in their third year, correct functional propositions (standardized) did not differ between simple and complex systems (.61 vs. .57, respectively), p = .25 (see Figure 3A and Table 3 for mean values).

We also tested whether the results would change when the direction of an arrow connecting two concepts was included in determining the accuracy of a functional proposition. The analysis revealed a significant main effect of complexity, $F(1, 56) = 89.1$, $p < .001$, $\eta^2 = .61$. As for the more liberal coding, correct functional propositions were generally more frequently used in concept maps explaining the simple compared to the complex system (.62 vs. .43, respectively). There was no significant main effect of year of training, $F(1, 56) = .1$, $p = .73$, $\eta^2 = .00$, but a significant interaction between year of training and complexity, $F(1, 56) = 30.2$, $p < .001$, $\eta^2 = .35$. For participants in their second year, correct functional propositions were more frequent in concept maps explaining the simple compared to the complex system (.70 vs. .32, respectively), $p < .001$. In contrast to when a more liberal coding protocol was applied, this effect also emerged for participants in their third year, albeit to a smaller extent (.48 for high vs. .62 for low complexity), $p = .001$. Thus, when the arrow direction was taken into account, correct functional propositions were more frequent for the simple systems compared to the complex system in both groups.

Intersection over Union

The analysis revealed a significant main effect of complexity, $F(1, 56) = 179.3$, $p < .001$, $\eta^2 = .76$. Thus, the structural alignment with the reference map was generally higher for the simple systems than for the complex system (.41 vs. .15, respectively). There was no significant main effect of year of training $F(1, 56) = .6$, $p = .45$, $\eta^2 = .01$, and no significant interaction between complexity and year of training $F(1, 56) = 0.1$, $p = .84$, $\eta^2 = .00$. Contrary to our expectations, participants did not follow the structure of the abstraction hierarchy more closely when they knew the complex system (see Figure 3B and Table 3 for mean values).

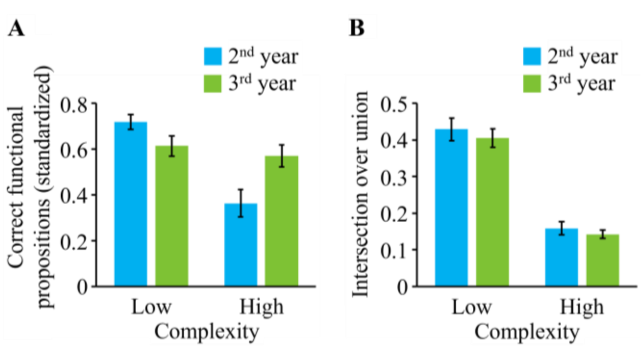

Figure 3: Comparison of low and high complexity maps regarding (A) content and (B) structure.

Discussion

Can mental models of system functioning be validly represented by concept maps that follow a functional abstraction hierarchy? In the present study, automotive and utility vehicle apprentices constructed concept maps for two simple systems (bike, traffic) and one complex system (fuel temperature control). All participants were expected to know the functioning of the simple systems. However, only one group (third year of training) had previously covered the complex system in class, whereas the other group (second year of training) had not. If the concept mapping method leads to valid assessments of participants' mental models, participants knowing the complex system should construct better concept maps of that system. For the simple systems, the method should not lead to any group differences. The expected pattern was observed for concept map content but not concept map structure.

Assessing Mental Model Content

The findings support the notion that concept maps focusing on means-ends relations can be used to assess participants' understanding of system functioning. Participants included more correct functional propositions in their concept maps when they had previously covered fuel temperature control in class compared to when they had not. Since the instructed method follows clear rules (i.e., connect concepts on adjacent levels of abstraction using means-ends relations), participants with a low understanding of system functioning might have also been able to construct good concept maps. However, this clearly was not the case. More specifically, participants who had not previously learned about fuel temperature control used more unspecified propositions and fewer functional propositions to explain system functioning. Misapplying the rules of the concept mapping method therefore does not seem to be an evident threat to the assessment of mental model content.

Assessing Mental Model Structure

To assess the structural quality of the concept maps (i.e., the similarity to the reference map following the abstraction hierarchy), we computed the intersection over union. This measure did not prove to be a valid indicator of participants' system understanding. While the structural similarity for the simple systems was high in both groups, it was low for the complex system in both groups. The first instinct might be to ascribe this finding to apprentices' lack of functional understanding. However, the group difference for correct functional propositions indicates that apprentices in their third year of training did understand the complex system's functioning better. In addition, the low quality of functional structure occurred for the complex, but not the simple systems. We will next discuss two other, more likely reasons for this finding, namely the requirements for constructing a coherent structure, and the appropriateness of the abstraction hierarchy for fault diagnosis in the automotive domain.

First, participants were provided with far more concepts to use for constructing the map about the complex system than for the maps about the simple systems (20/22 vs. 13). Previous research has shown that cognitive and meta-cognitive processes such as organizing and monitoring are required to construct a coherent map structure (Hilbert & Renkl, 2008). Since none of the participants were familiar with concept mapping following an abstraction hierarchy, it can be assumed that the structure-related requirements were high for all three concept mapping tasks. However, these requirements arise in addition to those posed by the content to be mapped (Correia & Aguiar, 2014). While participants in their third year of training have covered fuel temperature control in class, they likely have not achieved a level of proficiency that is comparable to that for the simple systems. This notion is supported by our findings regarding the comparison of two coding schemes. More specifically, participants in their third year of training had more accurate mental models of the simple compared to the complex system when a stricter coding scheme was applied. The added content-related requirements for the complex system therefore could have been too high for all participants. Further, this problem might have been enhanced by the rather short time allotted for the complex concept map (about 30 minutes). Since third-year apprentices were able to construct individual propositions that were functionally correct, they might have focused their resources on coping with the content-related instead of the structure-related demands.

Second, the specific structure of the abstraction hierarchy may not correspond to the structure of mental models that are useful for diagnosing faults in the car mechatronics domain. This potential mismatch already became apparent when constructing the reference map for fuel temperature control with two experts for fault diagnosis in automotive. According to both experts, a more appropriate structure of the concept map would have represented the physical connections between the system's components in the center instead of the bottom half of the hierarchy. Both experts would have included the means-ends relations to the respective functions and properties, but they would have chosen a different order for the abstraction levels in general. Additionally, they would have included relations representing physical connections (similar to a circuit diagram). The experts' focus on physical components and connections closely matches typical requirements during fault diagnosis in cars, namely identifying faulty components and replacing them (Schmidt & Müller, 2023). Another task in the same domain is car development, where decisions about the implementation of certain functions are typical. For this task, the original abstraction hierarchy representation might be more useful, and thus, the structure of a car developer's mental model might be more aligned with this representation. In sum, if the abstraction hierarchy structure does not correspond to the task requirements that people typically have to respond to, they likely have to transform their actual mental model structure into an abstraction hierarchy. Regarding the concept mapping method, our measure of concept map quality will reflect people's ability to conduct this transformation, not their actual mental model structure. Hence, suitable reference structures are necessary, and requirements from the domain and from typical tasks should be taken into account to identify them.

Limitations

Finally, some limitations of the present study should be kept in mind, namely the lack of balancing the order of conditions, the operationalization of system knowledge on the group level, and the lack of a performance measure.

Since the present study was conducted in a classroom setting and constructing the simple concept maps was part of the concept mapping training, the order of systems could not be balanced across participants. Thus, order and motivation effects cannot be ruled out. For example, it is possible that participants knowing the complex system would have been able to follow the abstraction hierarchy structure more closely if this map had been constructed earlier. However, given that the difficulties with the abstraction hierarchy also arose with domain experts, it is unlikely that they can fully be explained by effects of order and motivation.

Second, the two groups were based on the year of training, not on a more specific measure of participants' knowledge about fuel temperature control. If participants who knew the complex system's functioning were included in our sample of second-year apprentices, the obtained values might overestimate the mental model quality for people not knowing the system. Similarly, even though all participants in their third year of training have covered fuel temperature control in class, this does not guarantee that all of them understood its functioning. If participants who did not know the complex system's functioning were included in our sample of third-year apprentices, the obtained values might underestimate the mental model quality for people knowing the system. While this does not invalidate our general findings, it shows that our results cannot be used as a benchmark for the level of functional understanding.

Third, the present study did not relate the characteristics of participants' mental models to an external performance criterion, such as the ability to solve a problem occurring with fuel temperature control. Thus, we currently cannot say whether people who construct better concept maps are also better at solving problems within the system.

Conclusion

We examined whether concept maps with a focus on means-ends relations can be used to assess people's understanding of system functioning. While the content-related quality of concept maps corresponded to differences in knowledge about system functioning, the structural quality did not. One possible reason for this finding is a mismatch between domain requirements and the abstraction hierarchy structure used here. Identifying more appropriate structural references is an open task for future research.

Acknowledgements

Parts of this work were funded by the German Research Council (DFG, grant numbers AB 441/1-3, PA 1232/12-3).

References

Ade, N., Peres, S. C., Sasangohar, F., & Son, C. (2019). Analyzing procedure performance using abstraction hierarchy: Implications of designing procedures for high-risk process operations. In *Proceedings of the 22nd Annual International Process Safety Symposium*. College Station, Texas.

Bennett, K. B., Edman, C., Cravens, D., & Jackson, N. (2023). Decision support for flexible manufacturing systems: Application of the cognitive systems engineering and ecological interface design Approach. *Journal of Cognitive Engineering and Decision Making*, *17*(2), 99–119. https://doi.org/10.1177/15553434221118976

Bisantz, A. M., & Vicente, K. J. (1994). Making the abstraction hierarchy concrete. *International Journal of Human-Computer Studies*, *40*(1), 83–117. https://doi.org/10.1006/ijhc.1994.1005

Borst, C., Bijsterbosch, V. A., van Paassen, M. M., & Mulder, M. (2017). Ecological interface design: Supporting fault diagnosis of automated advice in a supervisory air traffic control task. *Cognition, Technology & Work*, *19*(4), 545–560. https://doi.org/10.1007/s10111-017-0438-y

Burns, C. M. (1999). Scanning patterns with ecological displays when abstraction levels are separated. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(3), 163–167. https://doi.org/10.1177/154193129904300307

Burns, C. M. (2000). Navigation strategies with ecological displays. *International Journal of Human-Computer Studies*, *52*(1), 111–129. https://doi.org/10.1006/ijhc.1999.0326

Burns, C. M., Bisantz, A. M., & Roth, E. M. (2004). Lessons from a comparison of work domain models: Representational choices and their implications. *Human Factors*, *46*(4), 711–727. https://doi.org/10.1518/hfes.46.4.711.56810

Bußwolder, P. (2015). The effect of a structured method on mental model accuracy and performance in a complex task. *Systems*, *3*(4), 264–286. https://doi.org/10.3390/systems3040264

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Correia, P. R. M., & Aguiar, J. G. de (2014). *Concept mapping informed by cognitive load theory: Implications for tasks involving learner-generated cmaps*. São Paulo.

Feuerstack, S., & Saager, M. (2022). Ecological interface design for efficient maritime traffic supervision. In *ICPS, Proceedings of the 33rd European Conference on Cognitive Ergonomics: Evaluating the reality-virtuality continuum*. New York, New York: Association for Computing Machinery. https://doi.org/10.1145/3552327.3552335

Florida Institute for Human & Machine Cognition (2019). CmapTools (Version 6.04) [Computer software].

Gary, M. S., & Wood, R. E. (2011). Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal*, *32*(6), 569–594. https://doi.org/10.1002/smj.899

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.), *Ablex series in computational sciences. Pathfinder associative networks: Studies in knowledge organization* (pp. 75–87). Norwood, New Jersey: Ablex.

Hall, T. J., Rudolph, J. W., & Cao, C. G. (2006). Fixation and attention allocation in anesthesiology crisis management: An abstraction hierarchy perspective. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(10), 1064–1067. https://doi.org/10.1177/154193120605001013

Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: What characterizes good and poor mappers? *Instructional Science*, *36*(1), 53–73. https://doi.org/10.1007/s11251-007-9022-9

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Jee, B. D., Uttal, D. H., Spiegel, A., & Diamond, J. (2015). Expert-novice differences in mental models of viruses, vaccines, and the causes of infectious disease. *Public Understanding of Science*, *24*(2), 241–256. https://doi.org/10.1177/0963662513496954

Johnson-Laird, P. N. (2013). Mental models and cognitive change. *Journal of Cognitive Psychology*, *25*(2), 131-138. https://doi.org/10.1080/20445911.2012.759935

Mumford, M. D., Hester, K. S., Robledo, I. C., Peterson, D. R., Day, E. A., Hougen, D. F., & Barrett, J. D. (2012). Mental models and creative problem-solving: The relationship of objective and subjective model attributes. *Creativity Research Journal*, *24*(4), 311–330. https://doi.org/10.1080/10400419.2012.730008

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York, NY: Elsevier Science Publishing Co. Inc.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*(3), 349–363. https://doi.org/10.1037/0033-2909.100.3.349

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, *38*(2), 260–278. https://doi.org/10.1002/1098-2736(200102)38:2<260::AID-TEA1005>3.0.CO;2-F

Ruiz-Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, *7*(2), 99–141. https://doi.org/10.1207/S15326977EA0702_2

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685

Schmidt, J., & Müller, R. (2023). Diagnosing faults in different technical systems: How requirements for diagnosticians can be revealed by comparing domain characteristics. *Machines*, *11*(12), 1045. https://doi.org/10.3390/machines11121045

St-Maurice, J. D., & Burns, C. M. (2017). Modeling patient treatment with medical records: An abstraction hierarchy to understand user competencies and needs. *JMIR Human Factors*, *4*(3), e16. https://doi.org/10.2196/humanfactors.6857