DIFFERENCES IN MENTAL MODEL DEVELOPMENT AMONG PSYCHOLOGY AND ENGINEERING STUDENTS OF A HUMAN FACTORS COURSE

Lawton Pybus, Allaire K. Welk, Douglas J. Gillan
North Carolina State University

How does a human factors practitioner's primary field of study affect the way he or she conceives of human factors concepts? Previous work has studied how mental models develop over the course of instruction, and how experts structure human factors knowledge. The present study longitudinally assessed mental models of human factors among students from psychology majors and students from engineering majors. Participants rated the relatedness of pairs of concepts for two units: one theoretical, and one applied. These data were used to produce Pathfinder networks for comparison. Results showed that students from the two majors held different mental models of the same concepts before and after instruction. Unexpected findings may indicate a possible application for mental model assessment: diagnosing issues in course design. Limitations, conclusions, and suggestions for future research are discussed.

INTRODUCTION

In everyday human factors applications, specialists from a variety of fields–for example, software development, information architecture, graphic design, marketing, or even psychology and engineering within human factors–must often collaborate to accomplish the goals of the team. These fields all have their specific foci, approaches, and biases. Outsiders to the field of human factors often have limited knowledge about how human factors theory and methods can be applied to improve systems and interaction, and the knowledge they do have may be misrepresented. Consequently, miscommunication and other team conflicts can and do happen.

How do human factors specialists structure their knowledge about the field, and how is that conceptual structure different from how outsiders think of human factors? Gillan, Breedin, and Cooke (1992) studied this by measuring two group's mental models of human-computer interaction (HCI): human factors experts and software development experts. Mental models are representations of the structure of concepts within a specific domain (Jonassen, 2009). They can be measured using a variety of knowledge elicitation and representation methods (Cooke, 1994; Rowe & Cooke, 1995, Chi, 2006). In knowledge elicitation methods, relationships between concepts within a domain are measured, either by word association, card-sorting tasks, or as in the case of Gillan and colleagues (1992), with a subjective numerical rating of the relatedness of terms (Fenker, 1975). Data from knowledge elicitation can then be represented. In relatedness ratings, after each pair of concepts has been rated, a proximity matrix is produced, which can then be represented spatially using multi-dimensional scaling (MDS; Kruskal, 1964), or Pathfinder networks (Schvaneveldt, Durso, & Dearholdt, 1989). In a Pathfinder network, a node represents each concept, and the links between nodes represent how related the concepts are perceived to be with each other. Concepts that are not perceived as very related are often only linked indirectly, while closely related concepts are typically directly linked. Gillan and colleagues found that human factors experts' and software development experts' networks were qualitatively different: human factors experts' networks were highly interrelated, while software development experts' networks had more central nodes and less interconnectedness.

If these differences are present, and may relate to problems in team performance, it would be useful to know how these differences develop. People naturally develop mental models of a domain as they learn about it, both individually and collectively (Jonassen, 2009). Therefore, one possible explanation for these observed differences may lie in instruction. Many researchers have studied the effect of instruction on mental model development, notably Goldsmith, Johnson, and Acton (1991), who found not only that student mental models tend to become more similar to their instructor's over the course of the semester, but that the degree to which student mental models converged with the instructor's was a strong predictor of classroom performance.

Although the effects of instruction are well studied, and other cognitive research has studied the relationship of prior knowledge in knowledge acquisition tasks (Pazzani, 1991), relatively little research has focused on how the prior knowledge that learners bring to instruction shapes mental model development. Fincher-Kiefer (1992) had participants with low, medium, and high knowledge of baseball read a radio broadcast of a baseball game and measured performance in making local and global inferences about the game. High-knowledge participants demonstrated better performance in making global inferences, and seemed to use a different overall strategy in understanding events. Although not measured directly, the author speculated that low and medium-knowledge participants had a developing or more flexible mental model of the subject, updating with new information, while high-knowledge participants' mental models helped to organize and give meaning to information. How might the prior knowledge of learners in human factors be related to their mental models' development, and the variety of knowledge structures previously observed? It may be the case that a student's major area of study would affect how they organize the information dispensed in a classroom. Instruction in human factors is now available at both the undergraduate and graduate level at a number of universities. Would a psychology major think differently about human factors than an engineering major? Typically, these courses, when offered, are upper-level courses for juniors and seniors, so students taking them would already have at least somewhat developed mental models of knowledge within their fields of study. If there are early differences among majors, the differences observed among experts might be present even among novices.

In the present study, we longitudinally assessed the mental models of psychology and engineering major students taking an introductory human factors course before and after instruction. Based on the previous discussion, we expect that psychology and engineering majors will initially have divergent mental models of domain concepts. We also expect that, following instruction, the mental models of the two majors will still differ, but will partially converge as they approximate the instructor's mental model.

METHOD

Participants

Forty-two students (22 males, M_{age} = 21.89 years, *SD* = 3.45 years) taking one of two introductory human factors course taught by separate instructors at a large university in the southeastern United States participated in assessments for partial course credit. Of these, 18 completed all three measurements, 18 completed two, and 6 completed only one. Fourteen students were in one section of the course, and the remaining 28 took the other. Thirty students were psychology majors, 6 were engineering majors, and the other 6, from other majors, were not included in the following analyses, for a total *N* of 36.

Measures

Participants' knowledge structures were elicited through relatedness ratings of every possible combination of pairs of seven terms in a half-matrix for two units of an introductory human factors course, for a total of 21 comparisons per unit. The first unit covered issues related to errors. The terms were: "slip," "mistake," "warning," "lapse," "violation," "error of omission," and "error of commission." The second unit covered human factors methods, and its terms were: "descriptive," "observation," "user goals," "usability testing," "task analysis," "heuristic analysis," and "system usability scale" (SUS). Thus one unit covered more *theoretical* concepts (errors), while the other was more *applied* (methods). Relatedness was rated on a scale of 0 to 10, with 0 indicating no relationship between the concepts, and 10 indicating a strong relationship.

Procedure

All assessments were taken online on the participant's computer using the Qualtrics web survey software. Prior to first assessment, participants completed demographic questions. Otherwise, after indicating whether or not he or she had participated in the previous assessment, participants were asked to rate the relatedness of the terms in a half-matrix, with these instructions: "Below is a half matrix with 7 terms on each axis. Please indicate the strength of the relationship between the two terms on a scale from 0 to 10. A rating of 0 would indicate that there is no relationship between the two terms, while a rating of 10 would indicate that the terms are strongly related For example, Solitaire and Excel could be rated a 0 because they are unrelated, Excel and Windows could be rated a 5 because they are both created by Microsoft, while Excel and spreadsheet could be rated a 10, because spreadsheets are used in Excel."

The assessments were taken at three time points: first, prior to any instruction on the first unit's concepts (Unit 1, Time 1); second, after all instruction on the first unit's concepts (Unit 1, Time 2), and prior to instruction on the second unit's concepts (Unit 2, Time 1); third, and finally, after all instruction on the second unit's concepts (Unit 2, Time 2). For the second assessment, participants completed two half-matrices of pairwise comparisons.

RESULTS

For all of the following analyses, the relatedness ratings of psychology majors and engineering majors were averaged to produce proximity matrix models for each group. We applied the Pathfinder algorithm to each model with parameters $r = \infty$ and q = n-1.

Mental Model Similarity

Unit 1: Theoretical Concepts. Prior to instruction, the mental models of the course concepts in Unit 1 were similar for psychology majors and engineering majors, as indicated by a significant similarity measure (S = .56, p = .010). After instruction, however, their mental models were different, as indicated by a non-significant similarity measure (S = .30, p = .290). Furthermore, while engineering majors' mental models were different before and after instruction (S = .36, p = .210), psychology majors' mental models were not (S = .71, $p <$ .001), possibly indicating that psychology majors had an earlier grasp of the material covered. Refer to Table 1 for all results from Unit 1.

Unit 2: Applied Concepts. Psychology majors and engineering majors differed in their mental models of the Unit 2 concepts prior to instruction (S = .09, p = .910). There were also significant differences before and after instruction for both psychology majors (S = .20, p = .580) and engineering majors (S = .33, p = .200). However, after instruction, the mental models of participants from either major were not statistically different (S = .50, p = .030), indicating convergence after instruction. See Table 2 for all results pertaining to Unit 2.

Qualitative Differences

Unit 1: Theoretical Concepts. See Figure 1 for representations of the two groups Pathfinder networks before and after instruction. Prior to instruction, engineering majors' networks were more interrelated in general, while psychology majors' networks had fewer direct linkages between concepts.

Engineering majors' networks had "violation" and "error of commissions" as central nodes. "Lapse," "slip," and "mistake" were related in the networks of students from each major, but psychology majors only saw "mistake" as related to "error of commission" while engineering majors thought all three were related to "error of commission."

Engineering major's networks differ between measurement time points, such that "error of commission," a central node prior to instruction, became a less directly linked node after instruction. "Warning" reversed in an opposite fashion, being more loosely related to the rest before instruction, but closely linked to central nodes after instruction. Psychology majors' networks differed only slightly. "Error of commission" mediated the linkage of "violation," "warning," and "mistake" prior to instruction, but is mediated instead by "errors of omission" afterwards.

After instruction, the networks of engineering and psychology majors were more similar. For example, "mistake" was a central node linked to "slip" in both networks. The engineering majors' networks remain somewhat more interrelated than psychology majors' networks.

Unit 2: Applied Concepts. See Figure 2 for representations of the two groups networks for Unit 2. Before instruction, psychology majors' networks had "usability testing" as a central node with almost all other concepts directly linked to it, while engineering majors' were less directly linked and had "task analysis" and "SUS" as central nodes. After instruction, "SUS" was replaced as a central node by "usability testing" for engineering majors. Furthermore, "observation," which was related to "task analysis" prior to instruction, was related to "usability testing" afterwards. For psychology majors, "task analysis" moved from an indirect link before instruction to a more central node afterwards.

Mental models of human factors methods were widely different prior to instruction, but appear to converge after instruction. "Task analysis" was a central node for both groups, as was "usability testing." Engineering majors tended to describe more direct linkages among concepts than did psychology majors.

DISCUSSION

Our hypothesis, that engineering and psychology majors would have and maintain different mental models of human factors concepts before and after instruction, was partially supported. Although quantitative differences were not observed before instruction in Unit 1, qualitative differences were. Similarity measures suggested that the two groups' networks were different after instruction, yet qualitative inspections suggest that, while different, they partially converged. These two approaches to analyses were less at odds for the second, more applied, unit. The two groups had different mental models at the beginning of instruction, but partially converged afterwards. Thus, it may be the case that the qualitatively different mental models human factors experts and experts from other domains have of the field are present even among novices.

These equivocal results for the two kinds of concepts (theoretical and applied) within the same domain may suggest that the differences observed between students of different majors are dependent upon the kind of material being studied. If engineering majors and psychology majors understand theoretical concepts similarly at first, but differently after instruction, perhaps they are organizing theoretical information differently. On the other hand, if they conceive of applied concepts differently at first, but understand them similarly later on, they may be shoe-horning unfamiliar concepts into pre-existing knowledge structures, but organizing them similarly after instruction.

Use as Assessment

Following Goldsmith, Johnson, and Acton's (1991) work correlating mental model convergence and classroom performance, measuring student mental models has often been used as a form of assessment. Further examination of some unexpected results from this study may reveal another, similar, use: diagnosing issues in instruction design. For students of either major in Unit 2, some terms' relationships with others are highly variable and apparently arbitrary (e.g. "heuristic analysis," and to a lesser extent, "descriptive"). Furthermore, it was somewhat surprising that "usability testing" was not the primary central node in the networks of both groups after instruction, as task analysis can be considered a usability testing method. We attribute these anomalous results to emphases in teaching. Likely, the instructors gave extra attention to task analysis in instruction, as it is an important human factors method (Gillan, 2012), and perhaps, comparatively less to heuristic evaluation and descriptive methods. Instructors may use this information to shift emphases and try to fill out gaps in students' knowledge for future semesters.

Limitations

The conclusions of this study should be considered with the following limitations. The analyses presented here are averaged by group. We may better understand differences with analyses of individual mental models. Although participants were distributed across two different courses taught by separate instructors, both courses were taught within the Department of Psychology. Instructors from other fields (Engineering, Computer Science, etc.) would likely have different mental models of the domain that they would impart to their students. Likely for the same reason, there were a limited number of engineering majors who enrolled in this course. Furthermore, instructor mental models were not analyzed for this study, so the degree to which participants' mental models after instruction were similar to the instructors is not known.

Conclusions

The variety of approaches to human factors can sometimes pose a problem for team performance. One possible cause of this problem is that team members from different disciplines have different conceptions of how the concepts within human factors are related. In this study, we have shown that these differences may be present even among novices early in instruction. Participants were largely advanced students who varied by major, but not by personal experience working in the field, or by the discipline of the instructor. Some structural, conceptual differences in their mental models of human factors concepts were evident both before and after instruction. This research expands what we know about how prior knowledge is related to the development of mental models before and after instruction, and also provides an important application for educators in the field. Assessing mental models may provide valuable information for improving instructional material. Future study should look at how these observed differences affect team performance, and if the effects are largely negative, how to ameliorate those differences. On the other hand, as Gillan and colleagues (1992) pointed out, differences may be positive, enabling a broader problem space, and more creativity within a team. They suggest team members should learn to more effectively communicate those differences. Understanding them, then, is the first step.

Table 1. Similarity statistic matrix and probabilities for mental models of unit 1 concepts.

		Time 1		Time 2	
		Eng.	Psy.	Eng.	Psy.
T1	Eng.	–			
	Psy.	0.56^{**}	–		
T2	Eng.	0.36	0.18	–	
	Psy.	0.56^{**}	0.71^{***}	0.3	–

Note: * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$.

Table 2. Similarity statistic matrix and probabilities for mental models of unit 2 concepts.

		Time 1		Time 2	
		Eng.	Psy.	Eng.	Psy.
T1	Eng.	–			
	Psy.	0.09	–		
T2	Eng.	0.33	0.50^{*}	–	
	Psy.	0.50^{*}	0.20	0.50^{*}	–

Note: * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$.

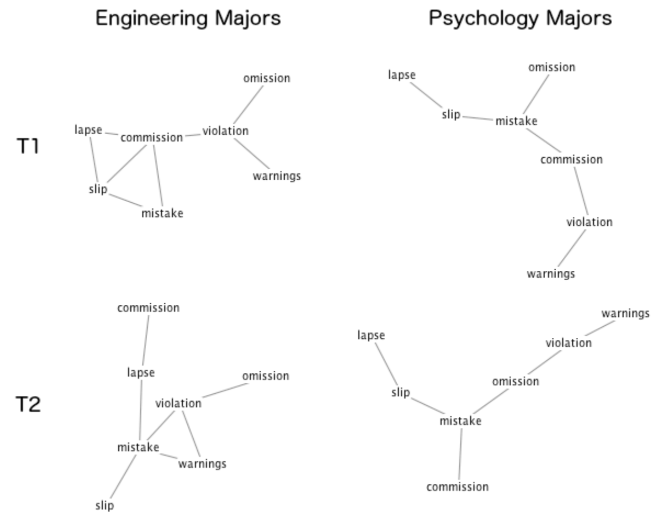

Figure 1. Time 1 and time 2 measurements of engineering and psychology majors' mental models of Unit 1: Theoretical Concepts.

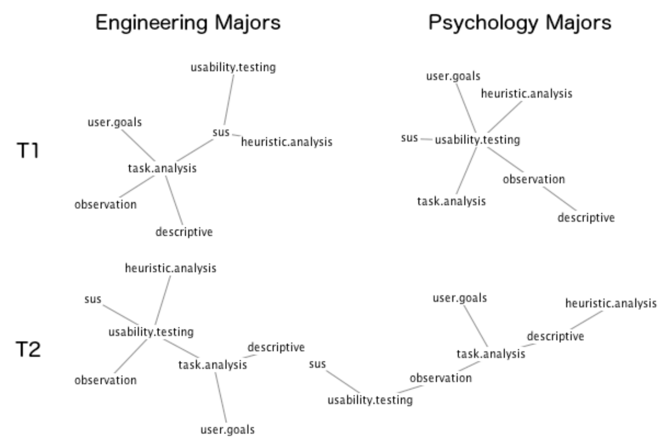

Figure 2. Time 1 and time 2 measurements of engineering and psychology majors' mental models of Unit 2: Applied Concepts.

REFERENCES

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, *41*(6), 801-849.

Chi, M. T. (2006). Laboratory methods for assessing experts' and novices' knowledge. *The Cambridge handbook of expertise and expert performance*, 167-184.

Fincher-Kiefer, R. (1992). The role of prior knowledge in inferential processing. *Journal of research in reading*, *15*(1), 12-27.

Gillan, D. J. (2012). Five questions concerning task analysis. *The handbook of work analysis: The methods, systems, applications, & science of work measurement in organizations*, 201-214.

Gillan, D. J., Breedin, S. D., & Cooke, N. J. (1992). Network and multidimensional representations of the declarative knowledge of human-computer interface design experts. *International Journal of Man-Machine Studies*, *36*(4), 587-615.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of educational psychology*, *83*(1), 88.

Fenker, R. M. (1975). The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures. *Instructional Science*, *4*(1), 33-57.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1-27.

Jonassen, D. H. (2009). Externally modeling mental models. In *Learning and Instructional Technologies for the 21st Century* (pp. 1-26). Springer US.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 416.

Rowe, A. L., & Cooke, N. J. (1995). Measuring mental models: Choosing the right tools for the job. *Human resource development quarterly*, *6*(3), 243-255.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. *The psychology of learning and motivation*, *24*, 249-284.