

Assessing Knowledge Acquisition in Academy Flight Screening (AFS)

Lt Abby Barger, Lt Kamille Kemp
Department of Behavioral Sciences & Leadership
USAF Academy, CO
kamille.kemp@wpafb.af.mil

Capt Chad Tossell
Department of Behavioral Sciences & Leadership
USAF Academy, CO
chad.tossell@usafa.edu

ABSTRACT

The Air Force relies on the Academy Flight Screening (AFS) program at the United States Air Force Academy (USAFA) to screen, motivate, and prepare pilot candidates for entry into Undergraduate Pilot Training (UPT). In this training, cadets are taught principles and techniques used in basic flying operations. In order to assess the effects of this training, we compare knowledge structures obtained from individuals that have completed AFS with knowledge structures obtained from individuals that have not graduated from AFS. Each subject was presented 435 pairs of basic aviation concepts and asked to rate their relatedness. These ratings were used to develop pathfinder networks and measures of internal consistency. Results show differences in cadets' understanding of important flying concepts. In addition, networks obtained from AFS graduates are more similar to one another than the non-graduates. Analysis of knowledge structures using pathfinder was also shown to be a useful adjunct to performance-based methods of assessing training already employed in AFS.

ABOUT THE AUTHORS

2Lt Abby K. Barger recently graduated from USAFA with a major in Human Systems Engineering. She will be attending undergraduate navigator training at Randolph Air Force Base.

2Lt Kamille M. Kemp recently graduated from USAFA with a major in Human Systems Engineering. She will be stationed at Wright-Patterson Air Force Base as a developmental engineer at Air Force Research Laboratory.

Capt Chad Tossell is an Instructor of Human Factors within the Behavioral Sciences & Leadership Department at the United States Air Force Academy. He directs the department's \$1.2M research technology capabilities for faculty and cadets and the Cadet Summer Research Program. Capt Tossell graduated from the University of California, Berkeley in 2003 and received his Masters of Science in Applied Psychology from Arizona State University.

Assessing Knowledge Acquisition in Academy Flight Screening (AFS)

Lt Abby Barger, Lt Kamille Kemp
Department of Behavioral Sciences & Leadership
USAF Academy, CO
kamille.kemp@wpafb.af.mil

Capt Chad Tossell
Department of Behavioral Sciences & Leadership
USAF Academy, CO
chad.tossell@usafa.edu

INTRODUCTION

Academy Flight Screening

The AFS program at the United States Air Force Academy aims to screen, motivate, and train potential pilot candidates. Most cadets participate in this program during the academic year. Consequently, students are also taking 12-18 hours of additional college credit, participating in military training, and involved with some form of organized athletic activity (some even at the intercollegiate level). The entire AFS course is 47.5 hours. It is almost equally divided between academic/ground school and flying training. Before a student is allowed to attend undergraduate pilot training (UPT), he or she must successfully complete the AFS program by demonstrating the ability to satisfactorily fly an aircraft and passing a written examination.

Assessment of Flight Training

The central goal of pilot training at any level within the USAF is to develop expertise in flight performance. Typical methods to assess the effectiveness of this training may include evaluations of changes in knowledge as a function of training and performance results as observed and rated by instructor pilots. In this project, we concentrated solely on assessing knowledge acquisition outside of the live training environment. Specifically, we chose to approach measuring the expertise of AFS graduates by examining how they conceptually organize various aspects of basic flight and comparing this knowledge structure with non-AFS graduates.

Researchers have utilized several methods for researching and defining knowledge structures. For example, multidimensional scaling (MDS) encompasses a variety of techniques that attempt to find a spatial layout of conceptual memory. MDS can represent large amounts of data in an easy to interpret format but it fails to capture asymmetrical and local relationships between concepts (McDonald and Schvanveldt, 1988). Alternatively, discrete models use graphs with nodes (concepts) and links (distances between concepts) to

describe the local networking of concepts in memory. Pathfinder is an example of this sort of technique.

Pathfinder is a tool that transforms judgments of the relationship between concepts made by subjects into a corresponding knowledge structure. Other labels that may also be attached to this output include mental model, concept map, cognitive structure, and network structure. Generally, knowledge structures reflect the user's understanding of a system and can be created spontaneously or taught through training. Creating a correct mental model through explicit training is advantageous because the students are then able to learn more effectively and adapt better to novel situations (Wickens & Hollands, 2000). The study of mental models has many applications especially in the realm of education and training. For instance, if novices learn not only the information that they need to know but also an effective way of organizing it, they will more quickly become experts. Our objective in this research is to determine if AFS students are developing a specific and effective mental model of basic flight concepts.

Multiple research efforts highlight the difference in knowledge structures of experts versus novices in various domains. Schvanveldt, Durso, Goldsmith, Breen, and Cooke (1985) addressed the differences between memory structures of expert and novice Air Force pilots. Cooke and Schvanveldt (1988) conducted two studies exploring differences in cognitive structures between novice and experts in the field of computer programming. One conclusion reached from these research efforts is that not only do experts possess more knowledge than novices, but knowledge that is shared by the two groups is organized differently in memory (Cooke & Schvanveldt, 1988).

The purpose of this research is to determine if there are measurable differences between the knowledge structures of AFS students and cadets who have not been through the AFS program. There are two hypotheses for this study. First, there will be a difference in knowledge structures of basic flight concepts between those cadets who have completed AFS (more experienced) and those who have not

(novices), indicating the training effectiveness of the AFS program. Second, the more experienced group will rate concepts more similarly than the novice group. This study also has the potential to allow the AFS program to track progress of students in a more objective way which will allow them to more effectively train future Air Force Pilots.

METHOD

Participants

A total of 36 USAFA cadets participated in this study, 18 of those had completed AFS and 18 had not. All participants were students enrolled in an Introductory Leadership Course and responded to a recruitment flyer. Demographic data was collected via a short survey at the beginning of the study. Subjects varied very little in flying experience and their objectives for completing AFS.

Materials

Two Dell Latitude D600 laptop computers were used to administer a survey via a graphical user interface created with MATLAB programming software. Pathfinder software, also MATLAB-based, was used for analysis of the data.

Procedure

Participants were recruited via the participant pool managed by the Behavioral Sciences & Leadership Department at the USAF Academy. On the recruitment flyer, every cadet indicated whether or not s/he had completed the AFS program. This information was used to sort the participants into either the non-AFS (novice) group or the AFS (more experienced) group. Appointments were scheduled individually through email. Upon arrival at the Behavioral Sciences Laboratory, participants were pre-briefed, informed of

Rate the relevance between the following two concepts:

ground_track **engine_stoppage**

0 - Not Relevant

1 - Of Little Relevance

2 - Of Some Relevance

3 - Moderately Relevant

4 - Very Relevant

5 - Extremely Relevant

Next

EXPLANATION OF THE RATING SCALE

0= The concepts have nothing to do with each other

1= The concepts have very little in common and are only slightly related

2=The concepts have one or two things in common but are still fairly separate

3=The concepts have some things in common and are moderately related to each other

4=The concepts have several things in common and are related

5=The concepts have a lot of things in common and are extremely related

Figure 1: Questionnaire Screen

the risks inherent in the study and asked to sign an informed consent document.

The participants were first asked to provide demographic data on items that could potentially affect their knowledge of the flight concepts presented during the study and/or their motivation to learn flight concepts. It asked whether they had taken AFS or not, whether they were planning on taking AFS, whether they had a private pilot's license and whether or not they had taken Aeronautical Engineering 315 (a core class that teaches basic aerodynamic principles). Upon completing the demographic data, participants interacted with the screen as shown in Figure 1. This was the questionnaire that asked users to rate the relevance between a pair of concepts. When all pairs had been rated, the survey program output the participant data into text files which could be read by Pathfinder software.

After completion of the demographic survey, each participant completed the basic flight concepts survey. There were 30 flight concepts (Table 1) presented pairwise for a total of 435 questions.

No pairs were repeated or swapped, for example, a

Table 1. Flight Concepts

traffic pattern	runway orientation
coordinated flight	rudder
turn and slip indicator	yaw
proper configuration	power on stall
flaps	pitch
trim	power off stall
boldface	V _r
emergency procedures	steep approach
forward slip	SFL
ground track	engine stoppage
airspeed	glide speed
aim point	checklist
landing	proper spacing
take off	breakout procedures
wind conditions	centerline

participant would not be asked to rate both 'trim vs. flaps' and 'flaps vs. trim'. For each pair, the participant was required to rank the relevance between the two concepts on a scale from 0-5, with 0 indicating no relevance between the concepts and 5 indicating maximum relevance between the concepts. A key for the scale appeared on the screen for reference.

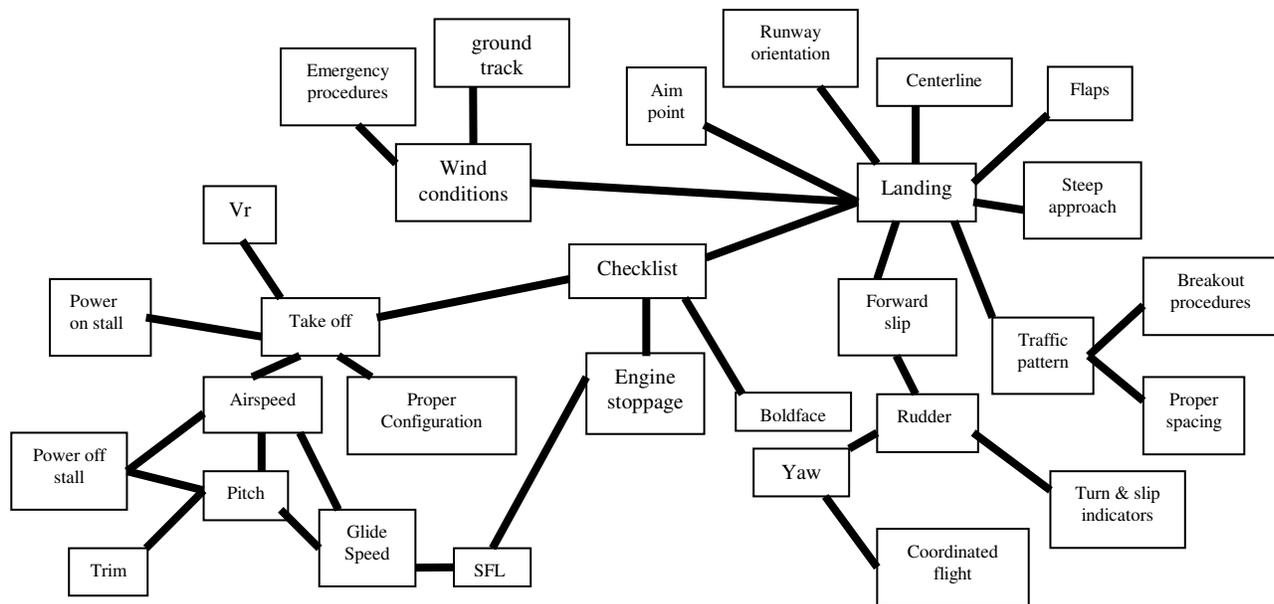


Figure 2. The Pathfinder Network Derived from the Average of AFS Graduate Ratings.

After completion of the survey, the facilitator ensured that the data file had properly saved each participant's responses to the survey. Participants were debriefed and given an extra credit receipt. The entire time spent in the laboratory ranged from 45 to 90 minutes.

All variables used in data analysis were based on the cadets' ratings of all the pairs of concepts. The resulting matrix after completion of survey contained 435 pairwise similarity measures representing proximities among all pairs of the 30 basic flight

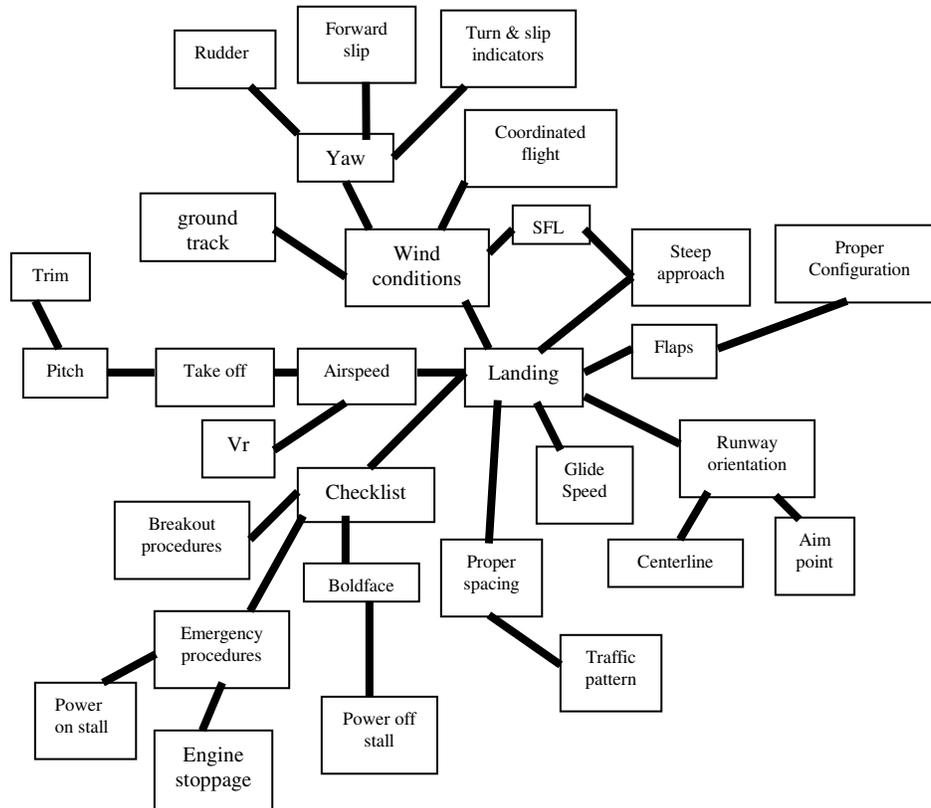


Figure 3. The Pathfinder Network Derived from the Average of Novice Ratings.

concepts. Such a matrix was obtained for each subject. The calculated psychological distances between the concepts were measures of their similarities to each other in terms of their relevance. The larger numbers are of greater similarity than the smaller numbers. The resulting matrix is a table in which both the rows and columns are the units of analysis and the cell entries are a measure of distance for each pair of cases. The difference between sets of qualitative elements (features) here forms the basis of a similarity model that can be used for fitting a network.

A network was derived from each set of ratings. Several additional variables were calculated. First, the data yielded coherence measures which assess the internal consistency of the ratings. Coherence is computed in two steps. First, a correlation of ratings for each pair of items is computed. For example, for

items 1 and 2, the ratings of item 1 across all of the other items is correlated with item 2 across all of the other items. This determines the extent to which the items in a pair are rated similarly against the other items. The second step correlates these correlations with the ratings given for the pairs. This can be seen as comparing the direct rating of a pair with the indirect relatedness inferred from the similarity of the ratings for the items in each pair. The coherence measure has been shown to reflect levels of expertise in that raters with more expertise in the concept domain generally produce higher coherence scores compared with less experienced raters. Coherence can also reveal unsystematic ratings that might be provided by a participant who does not take the rating task seriously resulting in careless ratings. Given that the rating task requires a large number of judgments, it is difficult for participants to be consistent by remembering earlier ratings. Rather, consistency

more likely stems from using a clear understanding of the domain as a basis of the ratings.

We also defined variables to assess the extent to which the networks obtained from individuals were similar to each other. Network similarity is assessed by first computing the ratio of the number of links two networks have in common over the number of unique links in both networks (i.e. the ratio of the cardinality of the intersection of the links in the two networks over the cardinality of the union). In a second step, the expected value of this ratio is subtracted from the ratio to yield the difference between the obtained value and the value expected by chance.

RESULTS

Pathfinder organized the data in a network consisting of 30 nodes representing the 30 basic flight concepts. The nodes were connected with weighted links based on the similarity data. Both the experienced (AFS graduates) and novice (non-AFS graduates) networks appear in figures 2 and 3. These connections represent the aggregated relatedness ratings between concepts as rated by both experienced and novice groups.

The novice structure was only a bit more complex (46 links) compared to the expert structure (33 links). Pathfinder also showed differences between experts and novices in terms of network similarity. This measure accounts for the number of shared links and intersections between individual networks taking into account network size and corrected for error. Two of the same networks would have a network similarity rating of 1. Two completely different networks (no similarities) would have a rating of 0.

In this study, both AFS graduates and novices had relatively low network similarity results. AFS graduates shared only 30% of their organization of basic flight concepts. Novices, on the other hand, shared an even fewer 20%. Thus, networks produced from the AFS graduate data ($M = .05$, $SD = .0117$) were significantly more similar to each other compared to the novices' networks ($M = .004$, $SD = .0113$), $t(14) = 5.86$, $p < .05$. Experienced respondents more consistently rated items in various combinations in the matrix and shared more characteristics in their conceptual structures compared to novices. Pathfinder's ability to classify experts and novices here is considered one

method to validate the precision of the network (Schvaneveldt et al, 1985).

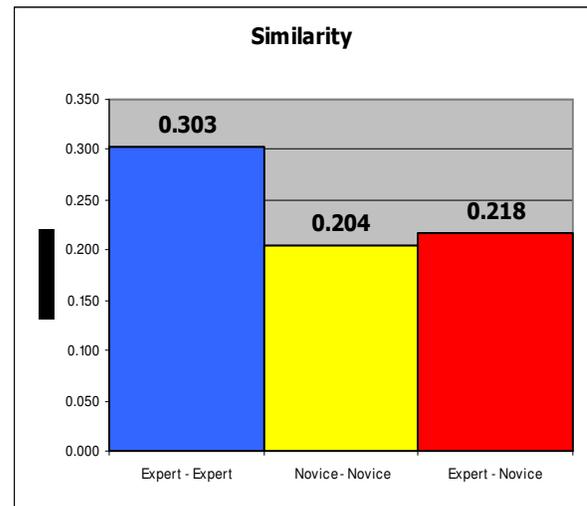


Figure 4. Network Similarity

Number of links in common is a measure of how similar two groups of networks are to each other. Figure 4 shows that the AFS networks had more links in common with each other than they had in common with the novices and than the novices has in common with the other novices.

The novices had about the same number of links in common with other novices as they did with those more experienced. The average novice network had only 35 links in common with the average expert network. The more experienced subjects on the other hand typically had approximately 45 links in common with each other. The probability that the more experienced networks would have this many links in common due merely to chance is .08%. Correcting for chance we find that the more experienced subjects still share approximately 23 links while the novices share only 12 and the two groups have about 13 in common. There is a significant difference between the number of links in common amongst the more experienced group as compared to the novice group $t(29) = 5.87$.

Identifying Discriminating Concepts

A final analysis was focused on identifying the concepts that are viewed most distinctly by experienced and novice pilots. Ratings of novice pilots were averaged. Similarly, ratings of AFS graduates were averaged as well. The differences

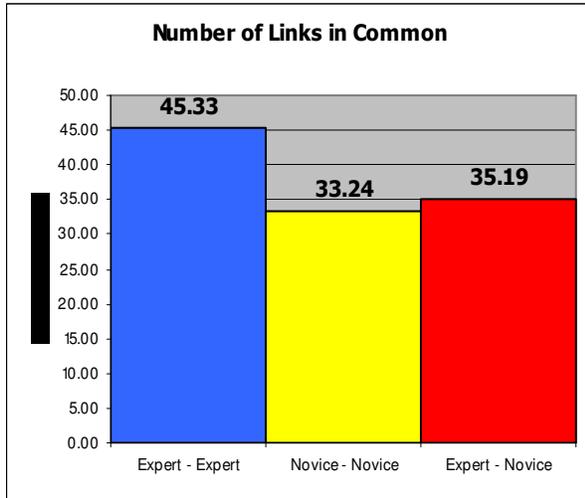


Figure 5: Number of Links in Common

between these two scores were computed. Summing the absolute values of these differences for each of the 30 concepts allows us to order them by the magnitude of the differences between experienced and novice ratings. This ordering is shown in Table 2.

Table 2. Concepts that Discriminate

Most Discriminating		Least Discriminating	
Rank	Concept	Rank	Concept
1	Emergency Procedures	1	Trim
2	SFL	2	Pitch
3	Proper configuration	3	Yaw
4	Power off stall	4	Turn & Slip Ind.
5	Coordinated flight	5	Rudder
		6	Boldface
		7	Runway Orientation
		8	Landing
		9	Aim Point

Knowing about the relative difficulty of these concepts could usefully feed back into the design of training by showing what the novices seem to know the least about. By ensuring that these concepts receive adequate coverage in the briefing and debriefing sessions, novices may be better prepared to deal with the corresponding aspects of their AFS flights.

In a qualitative comparison of the network diagrams produced by members of each group, there appears to

be only a slight difference. The simulated forced landing node in particular is telling. Our AFS group linked engine stoppage (the general cause of a forced landing) and glide speed (the most important consideration when making a forced landing) as most similar. Our novice group rated steep approach (a type of landing approach that may or may not be used in a forced landing) and wind conditions (an important consideration whenever landing an airplane) as most similar. We can see that although our novice group shows some understanding of basic aerodynamics they do not grasp some of the finer points of flying.

One potential reason for the lack of qualitative differences in networks could be that most cadets that participated in this study took an academic course on the fundamentals of aeronautics. It is possible that a set of more advanced concepts would reveal bigger differences. In addition, other measurement methods that focus on differences in procedural knowledge could be pursued as well.

DISCUSSION & CONCLUSIONS

Several aspects of the results reported above show that AFS training does produce a measurable difference in the way cadets conceptually organize basic flight information. In addition, there were overall increases in similarity among pilots after training. The increased similarity is apparently due to pilots being exposed to similar material/scenarios in training rather than to pilots learning more about members of their teams. Other studies of team cognition (see Cooke, Salas, Cannon-Bowers, & Stout, 2000) have often looked at increasing similarity among team members. The data reported here suggest that it may be useful to determine the extent to which increasing similarity is due to exposure to similar task conditions as opposed to training experiences with members of a team.

The results of our study could have some effects on the design of AFS training. The concepts we identified as leading to the greatest differences between novices and AFS graduates could receive more attention in the briefing and debriefing sessions accompanying the training sessions. Other ways in which the information we collect can be used in training are being considered. For example, students could be presented with networks like the one shown in Figure 2 to elicit discussion about the nature of the relationships depicted by the links. Inevitably such discussion draw out similarities and differences in the ways students think about the concepts. A hypothesis

that could be investigated is that training would benefit from such discussions, discussions that are driven by pilots underlying conception of flight.

Lastly, above results also show the sensitivity of pathfinder in capturing small differences in the representation of knowledge structures between slightly more experienced and novice individuals. AFS is primarily intended to screen potential flight candidates and only secondarily designed to provide basic flight training. Pathfinder's ability to provide quantitative distinctions in performance within the more experienced-novice paradigm here support future research into the utility of this tool as an adjunct to other performance measurement tools being developed by the Air Force Research Laboratories.

REFERENCES

- Brannick, M., Prince, C., & Salas, E. (2005). Can PC-Based Systems Enhance Teamwork in the Cockpit? In International Journal of Aviation Psychology (pp.173-187). Lawrence Erlbaum Associates, Inc.
- Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). *Measuring team knowledge*. *Human Factors*, 42, 151-173.
- Cooke, N. & Schvaneveldt, R. (1988). Effects of computer programming experience on network representations of abstract programming concepts. Int. J. Man-Machine Studies, 407-427.
- Gillan, D. J., & Schvaneveldt, R. W. (1999). Applying cognitive psychology: Bridging the gulf between basic research and cognitive artifacts. In F. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, M. Chi, & S. Lindsay (Eds.), *The handbook of applied cognition*. (pp. 3-31) Chichester: Wiley.
- Goldsmith, T. E., Johnson, P. J., Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88-96.
- Hinkle, D., Wiersma, W., & Jurs, S. (1994). Applied statistics for the behavioral sciences (3rd Ed.). Princeton, NJ: Houghton-Mifflin.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Prentice Hall.
- McDonald, J. & Schvaneveldt, R. (1988). The application of user knowledge to interface design. 289-338. In R. Guindon (Ed.), Cognitive Science and its Applications for Human-Computer Interaction (pp. 289-338). Hillsdale, NJ: Erlbaum.
- Rowe, A. L., Cooke, N. J., Hall, E. P. & Halgren, T. L. (1996). Toward an online assessment methodology: Building on the relationship between knowing and doing. *Journal of Experimental Psychology: Applied*, 2, 31-47.
- Schvaneveldt, R.W., Durso, D.T., Goldsmith, T.E., Breen, T.J, Cooke, N.M. (1985). Measuring the structure of expertise. Int. J. Man-Machine Studies, 699-728.
- Schvaneveldt, R., Durso, R., & Dearholt, D. (1989). Network structures in proximity data. The Psychology of Learning and Motivation, 249-284. Academic Press, Inc.
- Tossell, C. (2006). Representing the Structure of Expertise using Competency Constructs. Unpublished Thesis, Arizona State University.
- Wickens, C.D., & Hollands, J.G. (2000). Engineering Psychology and Human Performance. Upper Saddle River, NJ: Prentice-Hall Inc.