Acquisition and Use of Mental Operators: Multinomial Modeling versus ACT-R

Burkhard Müller (burkhard.mueller@psychol.uni-giessen.de)
University of Gießen, FB 06 Psychologie, Otto-Behaghel-Str. 10, 35394 Gießen

Wolfgang Schoppek (wolfgang.schoppek@uni-bayreuth.de)
University of Bayreuth, Department of Psychology, 95440 Bayreuth

Abstract

Knowledge about operators, about the conditions of their applicability, and about their effects is essential for effective interaction with the physical world. By combining two defining dimensions of this knowledge – abstractness of content and directionality of access – we can distinguish four classes of representational units: rules, instances, episodes, and structures. We present a multinomial model that measures the characteristics of these units. This model was applied to an experiment on the acquisition and use of alphabet-arithmetic operators (Müller & Gehrke, 2002). The multinomial model could be fitted very well to the data and allows calculating the proportions of the different kinds of mental operators. To compare these findings with a simulation of the specific cognitive processes, we developed an ACT-R model. Separating four cases of information processing in correspondence to the knowledge units in the multinomial model confirmed the estimates of the multinomial analysis. This finding demonstrates the usefulness of multinomial modeling as a statistical tool to investigate cognitive processes. Also, it provides converging evidence for the use of different kinds of knowledge, even in simple tasks.

Units of Mental Operators

Knowledge about causes and their effects is essential for our successful interaction with the physical world. More specifically, knowledge about operations, about the preconditions of their applicability and the resulting effects is crucial for successful planning. The term mental operator refers to this kind of knowledge throughout this paper. The content of a mental operator can be separated into three structurally different parts, namely the preconditions of applicability, the representation of the referred operation(s), and the effect or consequences of the application of the operator. Accordingly, three main types of knowledge use can be identified that require the specification of one of these parts if information about the other two parts is given: 1) Prognosis tasks, which require that the effect of operator application has to be predicted. An important instance of this kind of use is finding a way to solve a problem by forward chaining in problem space. 2) Retrognosis tasks, which require that the preconditions of operator application be identified. This kind of knowledge use is important if one tries to find a solution in problem space by backward chaining. 3) A third class of tasks requires the identification of an operation by comparing information about the initial state with information about the resulting effects. This kind of knowledge use is important if one has to select an appropriate action to achieve an intended effect. It is also involved in diagnosis tasks that require the identification of an operation as cause of an observed effect.

In the literature on problem solving and skill acquisition, different formats are proposed as representational units of mental operators. Four main classes of mental operators can be distinguished: production rules (e.g. Anderson & Lebiere, 1998), instances (e.g. Logan, 1988), episodes (= chunks in SOAR, Newell, 1990), and conceptual structures (e.g. Müller, 1999). These representational units can be classified by crossing directionality of access and abstractness of content as two essential characteristics of mental operators (see Table 1).

Table 1: Units of Mental Operators Ordered by Abstractness of Content and Directionality of Access.

	Access	
Content	Directional	Non-directional
Abstract, generalized	Production Rule (Anderson, 1993)	Structure (Müller, 1999)
Elementary, specific	Episode/ Chunk (Newell, 1990)	Instance (Logan, 1988)

Production rules in the ACT-R theory (Anderson & Lebiere, 1998) are intended to represent classes of situations. Thus, they are abstract in content. Access to these units is directional because it is limited to cues of the acquisition context that correspond to the elements of the if-part of a rule. Chunks in SOAR (Newell, 1990) are examples for directionally accessible units whose content is elementary. According to Newell, chunks represent the specific aspects of the acquisition context; access to these units is limited to the flow of information within this context. Examples for non-directional units of knowledge about operators are structures (Müller, 1999) and instances (for example, Logan, 1988). Müller proposed in his hypothesis of conceptual integration that structures represent the dependencies between initial and final state of an operation in an abstract way, access to these units is possible by cues that correspond to information stored in the structure. Halford, Bain, Mayberry, and Andrews (1998) have called this property omni-directional access. Instances differ from structures with respect to the abstractness dimension; they represent specific situations.

In this paper, we try to validate the results of a multinomial model that allows an estimation of proportions of the various knowledge types by comparing it with an ACT-R model of the same experiment. Since it is completely transparent what type of knowledge an ACT-R model uses for each problem, it is possible to compare the actual proportions of knowledge types used by the model with those estimated by the multinomial model.

An empirical study

Müller and Gehrke (2002) investigated the influence of knowledge use in the acquisition phase on the usability of knowledge about alphabet-arithmetic operators in a subsequent test phase. Participants practiced either the prognosis, the diagnosis, or the retrognosis task. In the test phase, participants had to answer all three kinds of tasks for old and new pairs of letters. Half of the participants in each training condition were administered a forced-choice test, the other half a verification test. For the present paper, only the forced choice test is relevant and will be considered in the following. Directionality of access should result in an advantage for the practiced task and a corresponding disadvantage for both other tasks. If the content of representation refers to concrete elements, performance should be better if old rather than new letter-pairs were involved.

Results showed that equal numbers of blocks were necessary for reaching the learning criterion. This finding suggests that the different uses of knowledge in the training tasks are equally difficult. However, response times were largest in the retrognosis task, suggesting difficulties to determine the initial state of an operation. The interaction of this effect with the day of practice indicated that participants can learn to apply operators in this way with equal efficiency. Relevant for the identification of the acquired representational units of knowledge, significant effects of use-specificity and familiarity were found. That means, (1) items of the practiced format were solved faster and more accurately than items of the non-practiced format, (2) old items were solved faster and more accurately than new items. On the basis of response times, these factors interacted. The effects were interpreted as evidence for significant proportions of directionally accessible as well as item-specific units of knowledge, i.e. significant proportions of rules and episodes as well as instances.

Measuring Units of Mental Operators

Müller and Gehrke (2002) interpreted their findings as evidence that mental operators are represented by various representational units. The proportions of these units, however, could not be determined on the basis of the statistical tools applied by the authors. Multinomial modeling provides a statistical methodology that allows estimating the probability of cognitive states. According to Riefer and Batchelder (1988), the linkage between (unobservable) mental states and (observable) behavior can be represented by a stochastic process: The probability of behavior C, denoted by p(C), can be defined as a function of a finite set of mental states T_1, T_2, …, T_n. This function becomes

$$p(C) = \sum_{i=1}^{n} P(C \mid T_i) P(T_i),$$

if one assumes that the n mental states are mutually exclusive and can cause the observable behavior C independently from each other.

Development of the Model

For the study of Müller and Gehrke (2002), a model is required that links states of mental operators and observable behavior for a two-alternative forced-choice test.

Behavioral categories are correct and wrong responses concerning four item classes: 1) old, congruent items, 2) old, non-congruent items, 3) new, congruent items, and 4) new, non-congruent items. Old items are test items that have been processed in the training phase, regardless of task format (prognosis, retrognosis, or diagnosis; e.g. "? +2 = C" is classified as old when "A +2 = ?" has been studied). New items are items that have not been practiced during training. Items are classified as congruent when the have the same task format, regardless of the specific content; otherwise they are classified as non-congruent (e.g. "? +2 = C" as test item is old and non-congruent, when "A +2 =?" was studied).

In a forced-choice test with two alternatives the realization of congruent and non-congruent items is straightforward. In a verification test, however, only parts of the information about the item must be presented for a short duration before the complete example is shown in order to distinguish congruent and non-congruent items. For congruent items, the part of information shown is the same as in the practiced items, for non-congruent items, different parts are shown.

Four parameters are required to represent the relevant mental states for the performance in the knowledge test: k, d, e, and g. Parameter k represents all kinds of knowledge, d represents the part of knowledge that is only directionally accessible, e represents the part of knowledge that is elementary, and g represents the state that a correct response is luckily guessed.

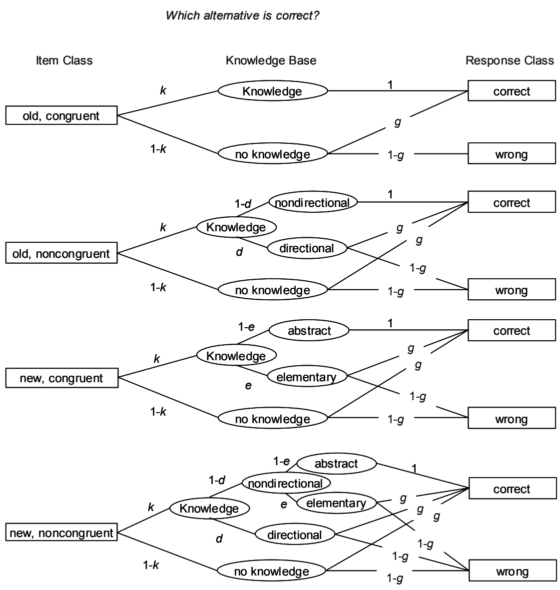

Figure 1: Multinomial Processing Tree Model for Responses in a Knowledge Test.

Figure 1 shows a decision tree for each of the four item classes that specify the relevant mental states / processes and their contribution to the observable performance in the respective task. The figure has to be read as follows: If test items contain information that has to be processed in the same way as in the training phase (old and congruent items), then all kinds of knowledge (k) will lead to systematically correct responses. If an item requires processing of old information in a non-practiced way (old and non-congruent items), then only the activation of non-directional knowledge (1-d) leads to a systematically correct response. If the test item requires processing of new information in a practiced way (new and congruent items), then only the activation of abstract (= non-elementary) knowledge (1-e) leads to systematically correct responses. If the test item requires processing of new information in a non-practiced way (new and non-congruent items) then only non-elementary *and* non-directional knowledge ((1-e) * (1-d)) will lead to systematically correct responses. In all other cases, responses will only randomly be correct.

The decision trees in Figure 1 can be described by equations (1) to (8) that express the probabilities of correct and wrong responses in a forced-choice test as function of the model parameters.

$$\mathrm{P}(\mathrm{correct}|\mathrm{Item}_{\mathrm{old,\ congruent}}) = k+(1\text{-}k)g \quad (1)$$

$$\mathrm{P}(\mathrm{wrong}|\mathrm{Item}_{\mathrm{old,\ congruent}}) = (1\text{-}k)(1\text{-}g) \quad (2)$$

$$\mathrm{P}(\mathrm{correct}|\mathrm{Item}_{\mathrm{old,\ non\text{-}congruent}}) = kdg+k(1\text{-}d)+(1\text{-}k)g \quad (3)$$

$$\mathrm{P}(\mathrm{wrong}|\mathrm{Item}_{\mathrm{old,\ non\text{-}congruent}}) = kd(1\text{-}g)+(1\text{-}k)(1\text{-}g) \quad (4)$$

$$\mathrm{P}(\mathrm{correct}|\mathrm{Item}_{\mathrm{new,\ congruent}}) = keg+k(1\text{-}e)+(1\text{-}k)g \quad (5)$$

$$\mathrm{P}(\mathrm{wrong}|\mathrm{Item}_{\mathrm{new,\ congruent}}) = ke(1\text{-}g)+(1\text{-}k)(1\text{-}g) \quad (6)$$

$$\mathrm{P}(\mathrm{correct}|\mathrm{Item}_{\mathrm{new,\ non\text{-}congruent}}) = k(1\text{-}d)(1\text{-}e)+k(1\text{-}d)eg+kdr+(1\text{-}k)g \quad (7)$$

$$\mathrm{P}(\mathrm{wrong}|\mathrm{Item}_{\mathrm{new,\ non\text{-}congruent}}) = k\,(1\text{-}d)\,e\,(1\text{-}g) + k\,d\,(1\text{-}g) + (1\text{-}k)(1\text{-}g) \quad (8)$$

The observable frequencies of the eight different response classes have four degrees of freedom. Thus, the four-parameter model has no degree of freedom left. To get a testable model, the parameter of randomly correct responses r can be restricted to .5, assuming no bias towards any response class. Fit of the multinomial model can be tested by the log-likelihood ratio of expected and observed frequencies, the resulting divergence statistic (G^2) is asymptotically χ^2-distributed. (for details of this calculation see Batchelder & Riefer, 1999).

Application of the model

Müller and Gehrke (2002) varied the presentation time of incomplete information within both test formats. The duration of this time was assumed to influence the importance of directionally accessible knowledge. If the acquired knowledge is only directionally accessible, then the time interval must be long enough for participants to compensate failed retrievals by additional processing. In the multinomial model, this influence means that the estimates of d should follow a monotone decreasing function of the presentation time of incomplete information. This influence was estimated separately for old (d_o) and for new items (d_n), because the training procedure should have led to high degrees of directionally accessible knowledge about practiced examples.

The model fits the data very well: $G^2(4) = 4.54$ (p = .34). The proportion of knowledge-based responses k is estimated very high. Moreover, a significant proportion of elementary knowledge (e) could be identified. No evidence for directionally accessible knowledge involving new items could be observed. The corresponding parameter d_n does not differ significantly from zero. As expected, the estimates for the corresponding parameter for old items (d_o) decreases with increasing time of presentation.

Given the assumptions underlying the multinomial model are valid, the proportion of the four kinds of knowledge units (see Table 1) can easily be calculated from the parameter estimates: The four kinds exhaustively represent the knowledge that contributes to performance in the test tasks; these units are mutually exclusive combinations of the two independent characteristics of mental operators. The proportion of production rules, for example, corresponds to the proportion of knowledge-based responses (k) where the knowledge is not elementary (1-e) *and* directionally accessible (d). Thus, the proportion of production rules is given by k * (1-e) * d_n. The proportion of the other kinds of knowledge formats can be calculated in an analogue way. Table 2 shows the results of this calculation.

Pres. Time*	Knowledge Unit			
	Structures	Rules	Instances	Episodes
1000	61.11	0.01	23.47	10.55
1500	61.11	0.01	26.08	7.95
2000	61.11	0.01	27.94	6.08

Note. * Presentation time of pre-information alone.

Table 2: Calculated Proportions of Knowledge Units in the Forced-Choice Test (%), grouped by Presentation Time of Incomplete Information

Structures are calculated as the largest proportion of units. Note that these estimates do not necessarily represent exclusively abstract and non-directional units of mental operators but may include pre-experimental knowledge (generalized procedures and perhaps item-specific associations).

An ACT-R model of Müller & Gehrke (2002)

We developed an ACT-R model to simulate the empirical findings of Müller and Gehrke (2002). Inspection of the knowledge that was used by the successful model should allow to validate the results of the multinomial model.

The model simulates three basic strategies of solving alphabet arithmetic problems, counting forward, generate and test, and retrieval of instances. As in other models of mental arithmetic (e.g. Lebiere, 1998), these procedures can be classified as calculation vs. retrieval strategies. In the early stages of learning, calculation strategies are more likely to produce correct results. For the three problem formats, retrieval strategies don't differ much, but there are format-specific versions of the calculation strategies. In prognosis problems with positive operators, the model simply counts up the alphabet the given number of steps. For minus-problems, we implemented a generate and test strategy, i.e. the model retrieves a letter some steps back, then counts forward and applies corrections, if necessary. When confronted with retrognosis problems, the model recodes the problem into a prognosis problem and solves it accordingly (e.g. ? +2 = M is recoded to M –2 = ?). With diagnosis problems the model makes an additional decision: One option is counting up from the first letter and checking if the second letter is within four steps up the first letter. If that's the case the model concludes it must be a positive operator, if not, it starts counting from the second letter and concludes it must be a negative operator. The second option is figuring out which of the two letters comes first in the alphabet and starting to count from that letter.

Unlike existing models of alphabet arithmetic (e.g. Johnson et al. 1998), which assume that retrieval is always tried first, our model instantiates more elaborate assumptions about the decision between strategies. First, the model makes use of the production parameter learning mechanism of ACT-R 5.0. Depending on the successes, failures and time for each strategy, ACT-R learns parameters for "probability of success", p, and "cost to reach a goal", c, and uses these parameters to calculate the "expected gain", $pg - c$, for all production rules (c is measured in terms of time, and g can be viewed as a constant). When more than one production rule is matching the content of working memory, the rule with the largest expected gain is selected. In early phases of learning, the expected gain of productions initiating retrievals is assumed to be lower than the expected gain of productions for calculation. With growing experience, the p values of productions for retrieval get higher, and, more importantly, the cost parameters c of productions for retrieval get much lower than those of productions for calculation. This gradual shift in p and c parameters results in an increased use of fact retrieval during the learning phase.

Since failed retrievals are very costly (one second with standard parameter setting) we introduced a "familiarity check" mechanism, which tries to judge from the presented information if a retrieval is likely to succeed. Such a mechanism was postulated by Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997), who also claimed that it would be very difficult to model in ACT-R. However, thanks to increased parallelism in ACT-R 5.0 it is now possible to model initiating a retrieval, waiting for a certain amount of time, and starting calculation when nothing has been retrieved within the waiting time (in our simulations, we set the waiting time to 450 ms).

A certain amount of speeding up calculation strategies with practice is produced by the production learning mechanism: from repeated retrievals of successor letters from declarative memory the model learns production rules that return the successor letter (and eventually even the successor number) directly.

We ran simulated experiments with 24 subjects each. Most parameters were set to their default values (baselevel learning: .5, retrieval threshold: .3) and we

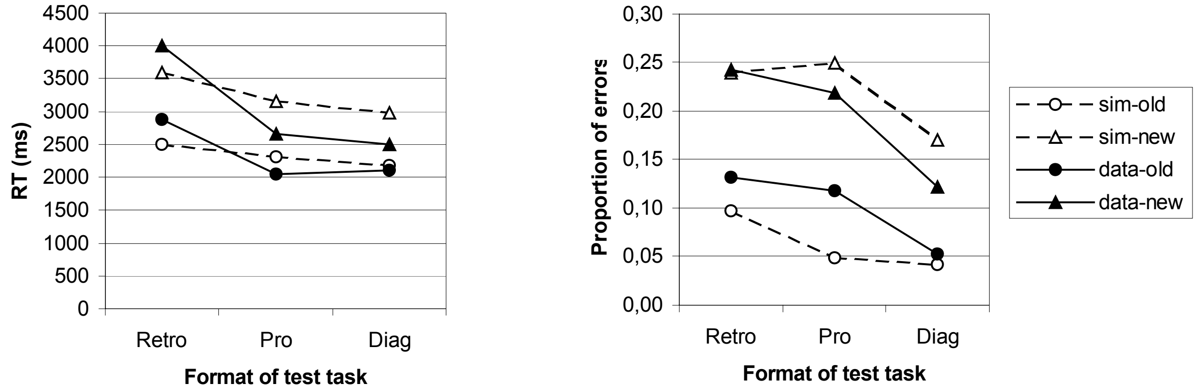

Figure 2. Means of response times and proportion of errors for old and new items in three problem formats in the test phase of Müller & Gehrke (2002).

didn't engage in extensive parameter fitting. As shown in Figure 2, the model reproduces the qualitative pattern of the data reasonably well. For response times, the model fits the six empirical data points with r^2 = .67, RMSD = 388. For errors, the fit is r^2 = .80, RMSD = .04.

The most striking difference is the underestimation of response times for retrognosis problems. The reason might be that recoding of problems into the prognosis format is more time consuming for human subjects than for the presented ACT-R model. The task-congruency effect (not represented in Figure 2) found in the data could not be reproduced with the model. We suppose this can be explained by the fact that the model starts with procedural knowledge of all possible strategies and only has to learn which one to use, whereas human subjects have to construct these strategies from general declarative knowledge (resulting in additional time for new problem formats in the test phase). Although considerable progress has been made in modelling the comprehension of instructions (Taatgen, 2001), modelling the formation of strategies from general knowledge is still a challenge for future developments that is worth being tackled.

To compare the multinomial analysis and the ACT-R model, it must be determined what kind of knowledge was used in the ACT-R model. Although only two types of knowledge are learned in the training phase – productions that automate counting through the alphabet and alphabet arithmetic facts – in the test phase we can distinguish four cases: (1) applying format specific facts (facts that were acquired in the context of the same problem format) – corresponding to episodes, (2) applying non format specific facts – corresponding to instances, (3) applying counting productions that were available before the experiment - corresponding to structures, and (4) applying counting productions that were acquired within the experiment – corresponding to rules. The proportions of these cases are determined and compared to the calculation of knowledge units based on the multinomial analysis. Table 3 shows the proportions of knowledge units determined by the inspection of the ACT-R model and those estimated by the multinomial analysis of the simulated data.

Table 3: Proportions of Knowledge Types (%) used by the ACT-R Model and Proportions of Knowledge Units Estimated by the Multinomial Analysis of the Simulated Data (%).

Model	Pres. Time*	Knowledge Unit Structures	Rules	Instances	Episodes
ACT-R	1000	58.5	0.0	29.0	12.5
	1500	59.9	0.0	28.6	11.5
	2000	64.2	0.0	23.3	12.5
MM	1000	48.6	0.0	37.3	3.8
	1500	52.1	0.0	30.8	0.0
	2000	63.9	3.4	27.7	2.9

Note. * Presentation time of pre-information alone.

The calculated proportions based on the multinomial analysis fit the reinterpreted proportions of knowledge used by the ACT-R model very well (r^2 = .92). The difference between these estimates and those reported in Table 2 can be attributed to the non-perfect simulation of the empirical data by the ACT-R model.

Conclusions

The multinomial model fitted the empirical data very well. In contrast to the statistical analysis of Müller and Gehrke (2002), the proportions of the four classes of knowledge units could be estimated. The ACT-R model fits the data to a reasonable degree. It was surprising that the model had difficulty reproducing the use-specificity effect. The multinomial analysis of the simulated data indicated that no abstract rules and only very few episodes were acquired and used by the ACT-R model. These calculations match the proportions of four different cases of information processing by the

ACT-R model very well. This finding supports the proposal of Batchelder and Riefer (1999) that multinomial modeling provides a promising way to combine the precision of very detailed process theories like ACT-R with the advantages of inferential statistics of rather general models like ANOVA.

From the other point of view, the ACT-R model has helped specifying the knowledge types postulated in the multinomial model. Structures could be reinterpreted as the use of generally applicable prior knowledge about the domain that needs not necessarily be abstract.

The distinction between abstractness of content and directionality of access as central aspects of mental operators differs from a recent suggestion of Pirolli and Wilson (1998) to use knowledge content and knowledge access as two characteristics to separate considerations at a knowledge level from those at a symbol level. The multinomial model of the present paper shall serve as a tool to integrate findings regarding representational characteristics that are proposed within different process theories of cognition. This integration should help specifying a detailed theory of mental operators at a process level. This level corresponds to the symbol level in the terminology of Pirolli and Wilson.

Future applications of the model can test the influence of important variables like instruction or types of practice on the characteristics of mental operators. Different theories can be tested by contrasting predictions concerning the change of parameters by these variables. Thus, the multinomial model shortens the chain of arguments and allows testing the influence on theoretically relevant entities directly.

Acknowledgments

This research was supported by a grant from the German Research Foundation (DFG: Mu-1333/4).

References

Anderson, J. R., Fincham, J., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 932-945.

Anderson, J. R. & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Lawrence Erlbaum Associates.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. Psychonomic Bulletin and Review, 6, 57-86.

Halford, G. S., Bain, J. D., Maybery, M. T., & Andrews, G. (1998). Induction of relational schemas: Common processes in reasoning and complex learning. Cognitive Psychology, 35, 201-245.

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. Psychometrika, 59, 21-47.

Johnson, T.R., Wang, H. & Zhang, J. (1998). Modeling speed-up and transfer of declarative and procedural knowledge. Proceedings of the twentieth annual meeting of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Müller, B. (1999). Use specificity of cognitive skills: Evidence for production rules? Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 191-207.

Müller, B. (2002). Single-use versus mixed-use learning of transformations: Evidence for conceptual integration. Experimental Psychology, 49, 45-56.

Müller, B. & Gehrke, J. (2002). Acquisition and use of mental operators: Effects of type of practice. Experimental Psychology, 49, 141-152.

Newell, A. (1990). Unified theories of cognition. Cambridge, Mass.: Harvard University Press.

Pirolli, P. & Wilson, M. (1998). A theory of the measurement of knowledge content, access, and learning. Psychological Review, 105, 58-82.

Rabinowitz, M. & Goldberg, N. (1995). Evaluating the structure process hypothesis. In F.E. Weinert & W. Schneider (eds.), Memory performance and competencies: Issues and growth in development. (pp. 225-242). Mahwah, NJ: Lawrence Erlbaum Associates.

Riefer, D. M. & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. Psychological Review, 95, 318-339.

Schunn, C.D., Reder, L.M., Nhouyvanisvong, A., Richards, D.R. & Stroffolino, P.J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 3-29.

Taatgen, N. A. (2001). A model of individual differences in learning air traffic control. In E.M. Altmann, A. Cleeremans, C.D. Schunn, & W.D. Gray (eds.), Fourth international conference on cognitive modeling. (pp. 211-216). Mahwah, NJ: Lawrence Erlbaum Associates