Feedback Effects in the Acquisition of a Hierarchical Skill

Andrew Corrigan-Halpern (ahalpe1@uic.edu)
Stellan Ohlsson (stellan@uic.edu)
UIC Department of Psychology, 1007 W. Harrison
Chicago, IL 60607

Abstract

Complex cognitive skills cannot be learned without feedback. There are reasons to believe that positive and negative feedback function differently. For skills that are represented hierarchically, feedback can provide information locally, concerning individual actions, or more globally, referring to higher level goals. A 2-by-2 factorial experiment showed an interaction between type of feedback (positive or negative) and scope of feedback (global or local). Contrary to the wide spread belief in the effectiveness of positive feedback, local negative was most effective. Global negative feedback was least effective. Positive feedback fell between these two conditions and was less affected by scope. The results are discussed in terms of the information content of the feedback.

Introduction

A large part of learning consists of gradually fine-tuning our efforts until we have succeeded in the task at hand. Learners may be able to evaluate progress without aid in simple tasks, but for complex tasks they rely on feedback (Kluger & DeNisi, 1996). For complex tasks, performance will often involve the execution of many goals, each involving the application of component skills (Kotovsky & Simon, 1973).

Many of the studies that employ feedback use it in the service of other research interests and are less concerned with comparing one feedback condition to another. Feedback is used to establish the levels of learning that the purpose of the experiment requires, but is not itself the object of study. Consequently, we know less about the cognitive function of feedback than one might expect. Studies that do focus on the feedback itself report contradictory findings and no current theory can fully explain why feedback is effective.

Some contrary findings may be explained by the fact that studies take one of two approaches, either investigating the informational or the motivational impact of feedback (Kluger & DeNisi, 1996). However, inconsistencies are not eliminated when we limit our survey to one or the other approach. In this article, we focus on the information aspect of feedback.

Background and Rationale

It seems obvious that if feedback is helpful, then more feedback should be more beneficial than less feedback. Many investigations have studied the impact of feedback rate (Salmoni, Schmidt, & Walter, 1984). In some cases, increasing the rate improves performance (Kulik & Kulik, 1988; Salmoni et al., 1984; Schmidt et al., 1989; Thorndike,

1927). In other studies, increasing the feedback rate is found to hinder performance (Bourne, 1957; Bourne & Bundeson, 1963; Schroth, 1997). This indicates that the effect of feedback is strongly mediated by other variables.

We would expect the content of the feedback itself to be one such variable. Studies in social psychology suggest that feedback is less effective when it draws attention to general performance or to self-efficacy and it is most effective when it draws attention to the task itself (Kluger & DeNisi, 1998). Another relevant variable is the type of task the learner is acquiring knowledge about. For complex tasks, feedback might foster more active processing, enhancing performance. For simple tasks, constant feedback may interfere with deeper evaluation, reducing performance. (Schmidt & Wulf, 1997). A third variable that affects the outcome reported in a study is the manner in which feedback is measured. The most effective feedback condition as measured during training might not be the most effective condition as measured by transfer measures (Schmidt, Young, Swinnen & Shapiro, 1989; Scroth, 1997).

In the study reported in this paper, we focus on yet another dimension of feedback: whether it provides information about appropriate, correct and useful actions, which we refer to as *positive feedback*, or whether it provides information about errors and mistakes, which we call *negative feedback*. Empirical work supports the notion that negative and positive feedback have qualitatively different effects (Taylor, 1991).

There is widespread belief that positive feedback is more effective than negative feedback. A simple argument about *information content* supports this belief: When a learner receives information that an action is appropriate, correct or useful, this requires neither change nor interpretation. The straightforward implication is that he or she should repeat that action when a similar situation arises in the future. When a learner receives information that his or her action was an error, he or she merely knows to avoid that particular action in the future, but not which action to perform instead. Positive feedback provides definite information, while negative feedback requires interpretation of the cause of the error and the selection of an alternative action. The interpretation process for negative feedback can be quite complex (Ohlsson, 1996).

The argument from information content is even stronger if we factor in the *scope* of the feedback. Many situations in real life do not provide feedback immediately after every elementary action. The effects of one's behavior might be delayed until after a series of actions has been completed

(Salmoni, Schmidt, & Walter, 1984). We call feedback that refers to a single action as being *local* in scope. Feedback that refers back to a series of actions is *global* in scope. The interpretation problem for global negative feedback is even more difficult than for local feedback. If a series of actions ends in an undesirable outcome, which part of the underlying cognitive skill should be affected by this feedback? If the sequence of actions is controlled by a subgoal, which in turn is dominated by a superordinate goal, which goal or goals should be affected by the negative feedback? During problem solving, a correct higher goal can dominate an incorrect lower goal, and vice versa. In prior work with a simulation model, we have shown that situations can arise in which a hierarchical, feedback driven system fails to learn because a superordinate and correct goal cannot recover from the negative feedback generated by incorrect lower level goals or actions (Corrigan-Halpern & Ohlsson, 2001). Ohlsson & Halpern, 1998). In contrast, global positive feedback does not seem to pose a more complex interpretation problem than local positive feedback. Learning that an entire sequence of actions was correct should be more powerful than receiving similar information about a single action.

However, empirical and theoretical research is not fully in accord with the implications of the argument from information content. Some researchers have indeed found that positive feedback is more effective than negative feedback (Greeno, 1974). However, in other cases, negative feedback was found to produce better performance (Mesch, Farh & Podsakoff, 1994). In prior work using a computer simulation model, we showed that some combinations of learning mechanisms imply that negative feedback will have a stronger impact on learning rate than positive feedback (Ohlsson & Jewett, 1997). The interaction with scope appears not to have been systematically investigated.

Closer scrutiny of the information content argument itself shows that it overlooks potential interactions with some of the factors mentioned above. The difference between practicing a simple task – drill – and solving an unfamiliar problem might be important. During problem solving, many actions are taken tentatively, with little or no rationale. When such an action fortuitously generates positive feedback, the learner only gains knowledge about how to solve that particular problem. To learn something that transfers to a related but different problem, he or she must figure out *why* the action worked. Hence, in this case, positive feedback requires interpretation. On the other hand, when the learner has acquired a rationale for his or her responses, then the positive feedback arrives after learning.

In summary, a straightforward application of the information content argument suggests that positive feedback should be more effective than negative feedback and that the advantage should be greater for large than for small scope, but this argument overlooks potential complicating factors. In the present study, we aimed to increase available information about these issues by systematically varying both feedback type and feedback scope, and by assessing the outcome during learning, after learning, and with transfer tasks.

Method

Participants. Ninety-four undergraduates participated in return for class credit.

Task. The subjects mastered a version of the sequence extrapolation task studied by, among others, Simon and Kotovsky (1963). The subject is shown a series of letters that follow a specifiable pattern (e.g., MABMCDM ...). Then he or she is asked to extrapolate the pattern to N additional places (EFMGHM...). We used a type of extrapolation problem that incorporated the hierarchical organization typical of the related problems studied by Restle (1970).

To make letter extrapolation into a task with multiple opportunities to receive feedback, we presented the given sequence via several short presentations and asked for a response after each one. The subjects viewed the given sequence for 20 seconds, then attempted to extrapolate it. They were asked to reproduce as much of the pattern as they could, guessing the letters for which they were uncertain. They received feedback on their extrapolation as described in detail below. Then the next trial (20 second study period, plus extrapolation attempt) began. The subjects went through 12 such trials.

To make the problem more difficult, a different sequence of letters was presented on each trial. For example, suppose a subject studied the sequence C A D F F D A C D B E G G E B D. He or she was prompted with the letter M and the correct extrapolation was M K N P P N K M N L O Q Q O L N. On the following trial, he or she saw a new instantiation of the pattern, e.g., F D G I I G D F G E H J J H E G. He or she was again prompted with M and the correct extrapolation was once again M K N P P N K M N L O Q Q O L N. The extrapolation prompt (M), and hence the correct extrapolation, was the same on each trial.

The tasks were presented on a 15 inch computer monitor with help of the PsyScope experimental control software. The given sequences were presented in black letters on a white background. When the 20 second study period ended, the given sequence was erased and the prompt M appeared on the screen to the left of a horizontal row of 15 answer boxes. The subjects gave their answer by clicking on the answer boxes in any order and typing in a letter. When done, they clicked on a 'Done' button, and the next trial commenced.

Design. In a 2-by-2 between-subjects design, we varied feedback with respect to type (positive vs. negative) and scope (local vs. global). *Negative feedback* consisted of the word "wrong" appearing in red text underneath an incorrect response. *Positive feedback* consisted of the word "correct" appearing in green text underneath a correct response. Each subject saw only one type of feedback.

Because feedback is intermittent in most real learning scenarios, we provided feedback probabilistically. A subject in a negative feedback condition received feedback on a random selection of 75% of his or her errors. For the remaining 25% of erroneous responses, plus all his or her correct responses, the subject saw the word "none" in white letters on black background below his or her response. The instructions emphasized that "none" meant that the response was *either* incorrect *or* correct. This feature prevented subjects from inferring that a response that did not receive negative feedback was correct (and vice versa). The presentation of positive feedback was analogous.

By *local feedback* we mean information about the correctness of a single letter. In contrast, *global feedback* referred to the natural chunks of the extrapolated sequence. Consider the given sequence CADFFDACDBEGGEBD. It consists of two parts, CADFFDAC and DBEGGEBD, which have the same structure but are separated by one position in the alphabet. The first part consists in turn of two subparts, CADF and FDAC, which are mirror images of each other; similarly for DBEG GEBD. Hence, the correct extrapolation MKNPPNKMNLOQQOLN consists of the four chunks MKNP, PNKM, NLOQ, and QOLN. In the global conditions, feedback was given with respect to these chunks by drawing a line underneath each group of four answer boxes; the word "correct", "wrong" or "none" appeared under the center of that line. The instructions emphasized that the feedback referred to the entire group of for letters. Thus, negative feedback meant that *at least one* of the responses in that group was wrong. In order to be get global positive feedback, all the responses in a chunk had to be correct.

In both the local and global conditions, the feedback remained on the screen for 45 seconds before the next trial commenced.

Procedure. The experiment began with an *Instruction* stage, in which subjects were given general information about the experiment. They were also taught the three relevant pattern construction operations (displace a letter or a group of letters one position forwards or backwards in the alphabet; repeat a letter or group of letters, and extend a sequence with its own reversal) with both verbal descriptions and examples. In the *Training* stage, the subjects went through twelve trials with respect to the target problem, as described above. In the *Assessment* stage, the subject solved the target problem twice without prior presentation of the given sequence and without feedback. That is, the prompt letter M appeared together with the answer boxes, and the subject attempted to fill them in; then M appeared again and the subject filled in the answer boxes once more. In the *Transfer* stage, the subject tried to solve a letter sequence problem with the same pattern as the pattern in the target problem. However, the prompt letter was T instead of M. In this case, the correct extrapolation consisted of a completely different sequence of letters. No feedback was given on the transfer problem.

Results

Analysis of Learning Stage

A Mixed ANOVA was performed to assess performance during the learning stage. All 12 Training trials were entered as a within-subjects factor. Feedback Scope (local or global) and feedback Type (positive or negative) were entered as between-subjects factor. The dependent measure was the number of letters correct per trial. There was a significant learning effect as shown in Figure 1, $F_{(990, 11)} = 42.21$, $p < .001$. There was no effect of feedback Scope, $F_{(90, 1)} = 1.33$, $p > .05$. There was no effect of feedback Type, $F_{(90, 1)} = 0.31$, $p > .05$. There was an interaction between Scope and Type, $F_{(90, 1)} = 8.45$, $p = .005$, indicating that the effectiveness of either feedback type is mediated by scope.
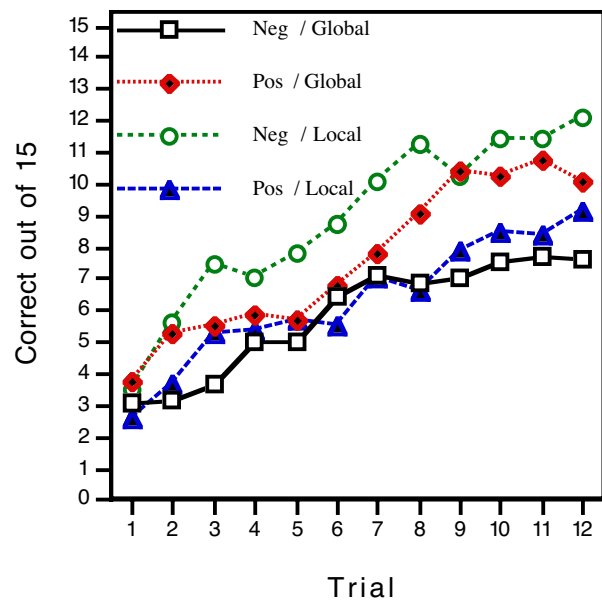

Figure 1. Performance on the training trials.

We hoped that by training subjects on the three construction rules, and by providing problems with a chunking structure, we would maximize the likelihood that they would represent the pattern hierarchically. To test whether this manipulation worked, we examined the inter response time for each position.

We computed solution times for each of the positions of the pattern. There was no time for the first letter of the pattern since this was given to subjects in the form of the prompt "M". We analyzed the last four trials of training, so that we could capture the final product of learning, the point where subjects would have come to represent the pattern hierarchically. Four subjects were removed from the analysis because they omitted responses on one or more of the trials. Figure 2 shows the result.

Subjects spent the longest times at the beginning of the pattern. The first three responses "K" "N" and "P" correspond to the first chunk. The high latency for the first 'K' probably reflected initial time planning to reproduce the pattern. The next four positions (5-8) correspond to the second chunk, "PNKM". Since subjects could reproduce these letters by reflecting the first chunk, response times were much quicker and there was a trend for these positions to form a horizontal line. The next chunk "NLOQ" involves the translation of the first chunk by one letter in the alphabet. This required more effort as indicated by the increase in latency from position 8 to position 9. The last chunk, "QOLN", could also be completed by reflection. Latencies match those of the second chunk and the horizontal trend is again present.
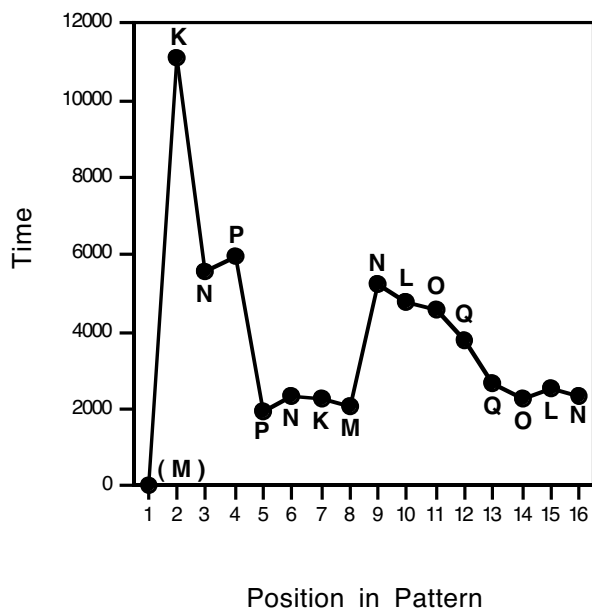

Figure 2. Time to complete each position averaged across the last 4 learning trials.

Analysis of Assessment and Transfer Stages

A mixed within-subjects ANOVA was performed for the Assessment trials. Both trials of the Assessment stage were entered as within-subjects measure. Type and Scope were entered as between-subjects measures. The dependent measure was the number of letters correct per trial.

There was no effect of feedback Type, $F (90 , 1) = 0.10$, $p > .05$ and no effect of feedback Scope $F (90 , 1) = 3.23$, $p > .05$. The interaction between Type and Scope was once again significant, $F (90 , 1) = 7.70$, $p < .01$, indicating that the effect of feedback type depends on scope. There were no effects involving the repeated measure. Figure 3 shows the results for the Assessment trials.

The two Transfer trials were entered as a within-subjects measure in an ANOVA. Feedback Type and Scope were entered as between-subjects measures. The dependent

measure was the number of letters correct per trial. There was neither an effect of Type nor of Scope, $F (90 , 1) = 0.05$, $F (90 , 0) = 2.15$, $p > .05$. The interaction between Type and Scope was once again significant, $F (90 , 1) = 7.58$, $p < .01$. Figure 4 shows the results for the Transfer trials.
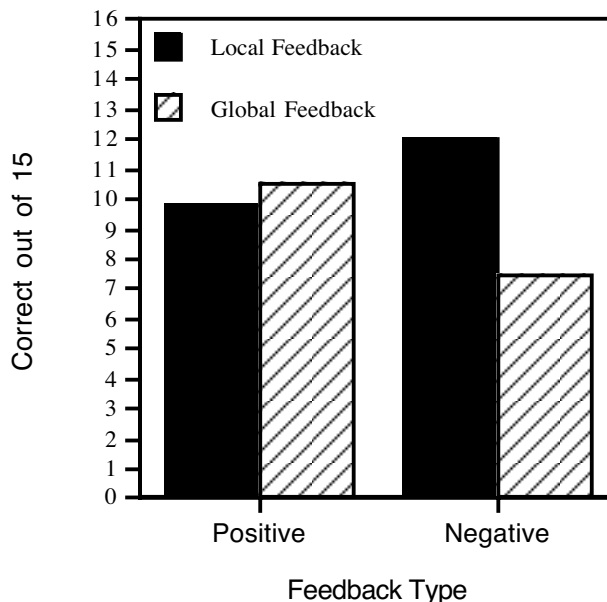

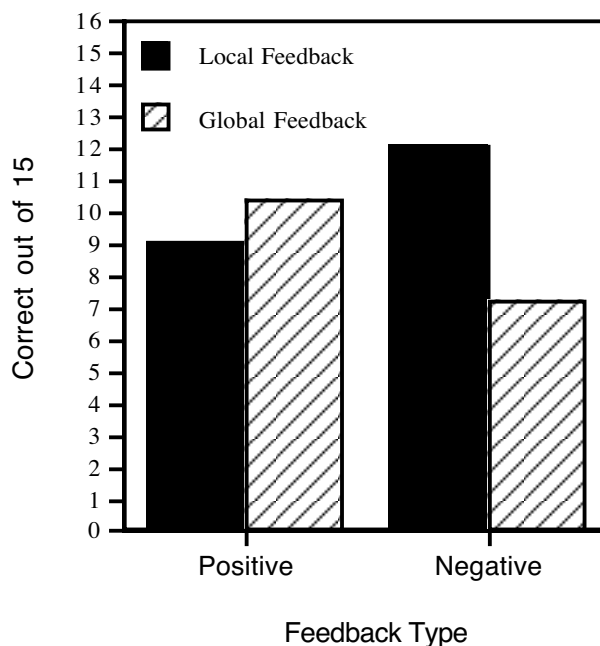Figure 3. Accuracy for the two Assessment trials.

Figure 4. Accuracy for the two Transfer trials.

In summary, we first showed that learning occurred during the Training stage of the experiment and we verified that subjects were representing the pattern hierarchically We showed that during all of the stages of the experiment there was an interaction between Scope and Type. This interaction is driven by the fact that negative feedback is maximized when provided locally, $F_{(90, 1)} = 9.88$, $p < .005$, while for positive feedback, there is no significant effect of Scope, $F_{(90, 1)} = 1.36$, $p > .05$.

Discussion

We presented data to show an interaction between feedback Type and feedback Scope. Local negative feedback was the most effective condition, while global negative feedback was the least effective one. Positive feedback fell between these two in effectiveness. Performance was similar for the two positive feedback conditions, suggesting that for this type of feedback Scope plays a limited role. This exact pattern of outcomes recurred during training (see Figure 1), in the assessment tasks (see Figure 3), and in the transfer tasks (see Figure 4).

This is not the pattern that is predicted by a straightforward formulation of the information content argument, since it predicts that the two positive feedback conditions should be best. According to the latter, negative feedback should provide less information regardless of scope, but this is not what we found. Furthermore, it is unclear what the information content argument predicts with respect to a transfer task in which the exact solution acquired during training is no longer sufficient. Nevertheless, the same pattern was observed in the transfer tasks.

We propose an alternative to the information content view, one where the informational content plays a role, but where it is not the only factor. In this study, negative local feedback was superior to negative global. Since negative local feedback provides direct information concerning individual responses, it is easier to interpret and hence can more directly be used to change subsequent responses. This explains why local negative feedback was more effective than global, but it does not explain our findings with respect to positive feedback.

Positive feedback applies in either of two cases. During problem solving, responses are sometimes made tentatively, without a rationale or reason. In the case of such fortuitously correct responses, positive feedback requires no less interpretation than negative feedback: *why* was the response correct? When the rationale has been worked out, the feedback arrives after learning has already occurred. The argument from information content ignores the dynamics of learning. Negative feedback is naturally received before learning occurs, and hence can influence and support learning. If the learner already knows how to generate a correct answer, the information in the resulting positive feedback is not novel and hence might not contribute strongly to learning.

This study has multiple limitations. The learning phase was short, only twelve trials. The feedback messages were limited to "correct" versus "incorrect" with none of the explanatory content that would be likely to accompany a feedback message in a realistic instructional situations (but not always in realistic situations where the feedback is 'delivered' by physical reality itself, e.g., in the form of a malfunctioning device). The transfer tasks were only moderately distant from the training tasks. The generalisability of the interaction between scope and type of feedback cannot be determined at this time. Future work will address these limitations.

What emerged clearly in this course of this study is that, contrary to the wide spread impression that feedback during problem solving practice is a topic that has been studied to death, we do not, in fact, know very much about the function of feedback. In particular, we do not understand the space of relevant parameters, and we do not know how currently identified parameters interact. It is plausible that the contradictory findings mentioned in the introduction are due to aggregation over overlooked interactions. A systematic experimental attack on the determinants of feedback effectiveness is warranted.

References

Bourne, L. E. (1957). Effects of delay of information feedback and task complexity on the identification of concepts. *Journal of Experimental Psychology, 54*, 201-207.

Bourne, L. E., & Bunderson, C. V. (1963). Effects of delay of informative feedback and length of postfeedback interval on concept identification. *Journal of Experimental Psychology, 65*, 1-5.

Chi, M. T. H., Feltovich, P. J., & Moser, J. M. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Corrigan-Halpern, A., & Ohlsson, S. (2001). Failure to Learn from Negative Feedback in a Hierarchical Adaptive System. In E. M. Altman & A. Cleermans & C. D. Schunn & W. D. Gray (Eds.), *Fourth International Conference on Computer Modling*. Fairfax, VA: Laurence Erlbaum.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Science, 6*, 270-292.

Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome feedback in multiple cue probability learning. *Organizational Behavior and Human Performance, 9*, 30-34.

Kluger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double edged sword. *Current Directions in Psychological Science, 7*(3), 67-72.

Kotovsky, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4(3)*, 399-424.

Kulik, J. A., & Kulik, C.-l. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research., 58*(1), 79-97.

Mesch, D. J., Farh, J.-L., & Podsakoff, P. M. (1994). Effects of feedback sign on group goal setting, strategies, and performance. *Group & Organization Management, 19(3)*, 309-333.

Ohlsson, S., (1996). Learning from performance errors. *Psychological Review*, *103*, 241-262.

Ohlsson, S., & Halpern, A. (1998). *Strength adjustment in hierarchical learning.* Paper presented at the 20th Annual Conference of the Cognitive Science Society, Madison, WI.

Ohlsson, S., & Jewett, J. J. (1997). Ideal adaptive agents and the learning curve. In J. Brzezinski & B. Krause & T. Maruszewski (Eds.), *Idealization VIII: Modeling in psychology*. Amsterdam, The Netherlands: Rodopi.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review, 77(6)*, 481-495.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin, 95*(3), 355-386.

Schmidt, R. A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors, 39(4)*, 509-525.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: support for the guidance hypothesis. *Journal of Experimental Psychology: Learning Memory and Cognition, 13*(2), 352-359.

Schroth, M. L. (1997). Effects of frequency of feedback on transfer in concept identification. *American Journal of Psychology, 110*(1), 71-79.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70(6)*, 534-546.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67-85.

Thorndike, E. L. (1927). The Law of effect. *American Journal of Psychology, 39*, 212-222.