



Effects of Feedback Interventions on Analogous Problem Solving: How do Structural Similarities Substitute Superficial Ones?

Agosto 2023, Vol. 15,
N°2, 60-77

revistas.unc.edu.ar/index.php/racc

Pulido, Marco ^{a*} , Morales, Valeria ^b , Ortega, Paola ^a , and Villicaña, Viviana ^a 

Original Article

Abstract

Resumen

Tabla de Contenido

This paper explored the effects of different feedback manipulations on analogous letter-string problem solving. Specifically, the present studies assessed the effects of the type of feedback (first experiment); long exposure to feedback and training (second experiment) and the number of feedback problems during the practice phase (experiment 3). Participants were college undergraduate students who voluntarily received and answered printed letter string problems. Results showed that correct answer feedback had a significant positive effect on deep structure problem solving; they also showed that problem solution occurs much faster and evolves in a more straightforward manner with this type of feedback. Results also suggest a direct relationship between the number of problems with correct answer feedback during the training phase and the number of correct answers during the testing trial. The discussion centers on the importance of experimental design as a tool for better understanding causal relationships in educational and other sciences.

Efectos de la retroalimentación sobre la solución de problemas análogos: ¿Cómo las similitudes estructurales substituyen a las superficiales? El estudio exploró el efecto de intervenciones de retroalimentación sobre la solución de problemas análogos de secuencias de letras. Se exploró el efecto del tipo de retroalimentación (primer experimento); exposición prolongada a la retroalimentación y entrenamiento (segundo experimento) y el número de problemas de práctica durante la práctica (tercer experimento). Participaron estudiantes universitarios que voluntariamente recibieron y contestaron los problemas. Los resultados mostraron que proporcionar la respuesta correcta tiene un efecto positivo y significativo en la comprensión de los problemas; también mostraron que la solución de los problemas es más rápida y menos variable, usando este tipo de retroalimentación. Los resultados también mostraron una relación directa entre el número de problemas en los que se recibió la respuesta correcta, y el número de respuestas correctas durante la fase de prueba. Se discuten los hallazgos en términos de la importancia del diseño experimental como herramienta para entender las relaciones causales en las ciencias de la educación, y en otras disciplinas científicas.

Introduction	60
Experiment 1	62
Experiment 2	65
Experiment 3	69
General Discussion	70
Aknowledgements	72
References	72
Appendix	76

Keywords: Feedback interventions, analogous problem solving, experimental design, college students

Palabras clave: Intervenciones de retroalimentación, solución de problemas análogos, diseño experimental, estudiantes universitarios.

Recibido el 13 de noviembre de 2020; Aceptado el 30 de diciembre de 2021

Editaron este artículo: Silvana Montes, Paula Abate, Sebastián Leonangeli y Macarena Fernández

The beneficial effects of feedback on learning and problem solving have a long history within psychology. First studied scientifically by Thorndike (1903), it gradually separated itself from its behavioral meaning (reinforcement) when it became clear that its key element is not exclusively that of an operant conditioning process. Instead, feedback also enhances learning

because it conveys information regarding the task. According to Sadler (1989, p. 82), “feedback helps the individual fill the gap between what is understood and what is aimed to be understood.” This is principally accomplished by providing the individual with information (Winne & Butler, 1994, p. 5740) that helps the learner “confirm, add to, overwrite, tune or restructure information in

^a Universidad Anáhuac, Facultad de Psicología, Ciudad de México, México.

^b Universidad Nacional Autónoma de México, Facultad de Psicología, Ciudad de México, México.

*Enviar correspondencia a: Pulido, M. E-mail: mpulidor100@gmail.com

Citar este artículo como: Pulido, M., Morales, V., Ortega, P., Villicaña, V. (2023). Effects of Feedback Interventions on Analogous Problem Solving: How do Structural Similarities Substitute Superficial Ones?. *Revista Argentina de Ciencias del Comportamiento*, 15(2), 60-77.

memory". Some scientists have suggested that the rapidly evolving challenges for education of the XXI century require instructional tools that quickly help students develop problem-solving capacities in a vast number of new areas of expertise (Reimers & Chung, 2016; Van der Kleij et al., 2012). Scientists have also suggested that Feedback Interventions (FI) may be an ideal strategy to meet these challenges because they are easy to adapt to computer-based learning environments (Corbalan et al., 2010; Murphy, 2010). FI have been used successfully within the domains of professional development (Toader & Lungu, 2015), early literacy education (Patchan & Puranik, 2016), vocabulary acquisition (Mohamed, 2020), algebra skills development (Bokhove & Drijvers, 2012), amongst many other relevant fields.

FI may be useful learning tool; however, the basic mechanisms that have such powerful effects on skill acquisition are still poorly understood (Catania, 1998; Johnson, 2013; Peterson, 1982; Weatherly & Malott, 2008). A vast number of issues remain unresolved (Mangiapanello & Hemmes, 2015); however, one objective of the present paper is to address issues regarding experimental design. A first one has to do with the way the dependent variable (DV) is measured. Specifically, pretest-posttest are prevalent (Kluger & DeNisi, 1996; Riper et al., 2009; Wisniewski et al., 2020), and this kind of measurement obscures the specific moments where feedback/reinforcement contingencies produce behavior change. Pretest-posttest designs are also blind to most changes in the learning curve; an incredible omission in any skill acquisition experiment (Anzanello & Fogliatto, 2011; Iversen & Lattal, 1991; Keller & Schoenfeld, 1950; Sidman, 1960; Skinner, 1938). A second issue regarding experimental design in FI, is that it is rarely possible to observe the control groups that would allow the reader to interpret the effects of feedback unambiguously (see for instance Crooks, 1988; Diehl & Serman, 1995; Ibarra-Sáiz et al., 2020; Kulhavy & Stock, 1989). Accepting the idea that one component of FI is operant conditioning (Carpenter & Vul, 2011; Grünke et al., 2017; King et al., 2000), assessing the effects of FI requires more than a simple experimental group and control group design (which is the most frequent practice, see for instance Van der Kleij et al., 2011; Kluger & DeNisi, 1996; Shute, 2008).

Assessing the effects of reinforcement on learning requires an experimental design that includes reinforcement and non-reinforcement conditions, however, as non-contingent feedback/reinforcement may be associated with behavior through accidental pairings (Kelley et al., 2017; Skinner, 1948; Zeiler, 1972), a random reinforcement condition is required to unambiguously assess the effects of FI. In brief, in the opinion of the authors (and others) assessing the effects of FI on skill acquisition requires repeated measures designs that allow a continuous analysis of the effects of the independent variables (IV) on the learning curve. It also needs a control group that allows the experimenter to differentiate between the effects of contingent and non-contingent response-feedback/reinforcement pairings.

Now that the reader is "warned" about the specific methodological approach of this study, it is time to turn to its specific objectives. Most research regarding FI has focused on the subject's ability to perform specific tasks. This pragmatic approach is understandable within the field of applied psychology; however, it has overshadowed an odd and troublesome issue regarding problem solving development. Specifically, using analogous problem examples to favor the independent solving of target ones (a frequent teaching practice), is frequently dependent on superficial similarities, rather than structural ones (Alfieri et al., 2013; Chi & VanLehn, 2012; Pulido, Almaraz, et al., 2010; Ross & Kilbane, 1997). There is considerable debate regarding how to identify superficial similarities (Goldwater & Jamrozik, 2019; Lobato, 2006; Schwartz et al., 2011); however, one may understand them as "those features of a problem that are not essential to its solution" (the problem context). In a similar vein, structural (deep) similarities can be understood as "those features that are essential for the solution of the problem." Based on this definition, a specific objective of the present research is to assess the effects of FI on college students' ability to solve problems using structural similarities. The authors hope that the experimental design will allow a fine-grained analysis of when and how structural similarities substitute superficial ones. Additionally, it is expected that non-contingent feedback/reinforcement control groups will allow an unambiguous interpretation of the results.

Regarding the DV of the study, the authors have chosen letter-string problem solution (LSP). These problems have been used for some time to study analogous problem solving (Burns, 1996; Pulido, Almaraz, et al., 2010). In a LSP, a determined alphabetical sequence is changed deliberately (for instance ABC is changed to ABD); after studying the first sequence, the subject is asked to modify a new letter-string in a similar way (for instance MNO). The resulting problem is frequently represented in the following way (ABC:ABD/MNO?); an MNP answer may be considered “correct”, (because in both problems, the last letter is substituted for the following letter in the alphabet; although other answers may also be considered correct for different reasons). LSP are appealing to scientists because generating different problem types is easy (combinations and permutations of all alphabet letters, gives a virtually infinite result). Additionally, the superficial vs. structural similarities phenomenon has been replicated in an important number of studies (Burns, 1996; Pulido, 2002; Pulido, de la Gama Valenzuela, et al., 2010). For instance, using a between-groups design, Pulido et al., (2005) observed that undergraduate students correctly solved letter-strings when both practice and target problems were in the same alphabetical order (for instance ABC:ABD/MNO?). However, when the alphabetical order of the problems was different (for instance ABC:ABD/OMN) the number of correct answers dropped dramatically. This finding was later replicated by Pulido, Hernández, et al., (2010) using a within-subjects, repeated measures design. Taken together, these findings suggest that when superficial similarities disappear, individuals have problems solving LSP. Two final considerations led the authors to choose this particular DV. First, in agreement with Jonassen’s (1997) classification, LSP fall within the category of ill-structured problems, this means that there may be more than one way to produce a “correct response”. This in turn allowed the present authors to arbitrarily select a “correct solution” and more clearly assess the potency of feedback contingencies to “shape” this specific response. Finally, also in agreement with Jonassen’s analysis, ill-structured problems are more interesting than well-structured ones because “they may be better suited for developing critical thinking,” from many perspectives a desirable problem trait (see Snyder, L. & Snyder, 2008 for a

review).

Experiment 1

One of the most common findings produced by the FI literature is that more elaborate feedback is associated with more dramatic differences between pretest and posttest conditions (Van der Kleij et al., 2015; Wisniewski et al., 2020). Based on this premise it was hypothesized that more elaborate feedback would likewise be associated with higher percentages of LSP being solved by structural similarities. Using the experimental designs and control groups described in the previous section, the first study assessed this possibility. The first study assessed the effects of either: a) no feedback, b) dichotomous feedback (the response is correct or incorrect), c) correct answer feedback (the individual is told, not only if the response is correct or incorrect, he/she is also provided with the correct answer). Finally, to rule out “superstitious” conditioning, some individuals received a non-contingent feedback condition. It was hypothesized that problem solving by structural similarities would be higher in the correct answer condition, relatively lower in the dichotomous feedback condition (and undistinguishable low in the non-contingent and no feedback conditions). To explore the DV in a systematic way, independent experimental groups could have learning, and practice problems in direct alphabetical order, inverse alphabetical order, or direct and inverse alphabetical order.

Methods

Participants

A total of 120 undergraduate college students from Mexico City participated in the study. Average age was estimated at 21.2 years with a SD of 1.2 years. The sample was predominantly male (54%). Most students lived with their families (87.5%). The Ethics Committee of the University approved the study. Only those students that read, signed, and accepted the terms of the informed consent form participated in the study. In order to participate, the individuals had to be registered (at the time of the study), in a bachelor degree program from the university.

Regarding sample size and statistical power, the authors took into consideration the experimental design. This study may be considered a between groups experiment with repeated measures for each subject. As such, the

importance of group comparability diminishes (because there are multiple sampling processes for each individual), and thus sample size becomes a lesser issue (Kirk, 1995; Ryan, 2013; Sidman, 1960; Skinner, 1938). Based on the previous argument, the reader may find the studies in this paper relatively undersized (when compared with other studies conducted within the FI literature). This is not atypical in repeated measures studies.

Instruments

All participants received a printed booklet with fifteen LSP. The subjects received one of three different family problems (the complete problems are presented in Appendix 1). In the first family, both learning and target problems were presented in direct alphabetical order (ABC:ABD/MNO?). In the second family both learning and target problems were presented in inverse alphabetical order (ONM:PNM/GFE?). In yet another family, the learning problem was presented in direct alphabetical order and the target problem in inverse alphabetical order (ABC:ABD/GFE?). Participants also received different types of feedback. In a control group, individuals were asked to solve the problems without any form of feedback (No Feedback condition). In a second control group (Non-Contingent or "Random" Feedback condition), the students solved the problems and were told that the problem was correctly or incorrectly solved. Feedback, however, did not depend on the subject's performance, it depended instead on a coin toss (coin tosses were prepared before working with the subjects). In the experimental groups, feedback was dependent on performance. In the first group (Dichotomous Feedback condition), subjects were told if their answers were either, correctly, or incorrectly solved. The second experimental group was essentially the same as the first one, however, when the subject produced an incorrect answer, he/she was presented with a printed card that showed the correct answer (Correct Feedback condition).

In brief, the experiment may be conceptualized as a 4 x 3 factorial design, the first IV was problem type, and the second was feedback constraints. Ten subjects were randomly assigned to each one of the twelve different experimental conditions. Each condition required the participant to solve 15 LSP.

Procedure

Experimenters approached individuals sitting on the library benches of the university and introduced themselves. They invited the students to participate in a "problem solving study," the students were told that participation was voluntary, and that the decision not to participate, would have no consequences (academic or otherwise). If the individual agreed to participate and signed the informed consent letter, he/she was randomly assigned to one of the twelve conditions. Subjects were presented with an example of an LSP; after that, they were given a booklet, and told to solve the problems, one by one, (and in the order established by the booklet). The experimenters registered the answers. The feedback provided for the participant depended on the condition to which he/she was assigned. Experiments were conducted in the library cubicles. The cubicles consisted of a table and two chairs; they were soundproofed and could be locked from the inside.

Results

Figure 1 shows the cumulative number of correct responses, for all subjects and conditions.

Figure 1 shows that in those conditions where model and target problem were presented in the same alphabetical order, the number of correct answers is invariably high, (and look very similar between themselves). However, in those conditions where model and target problems were presented in a different alphabetical order, the number of correct responses is comparatively lower, (except in the "correct answer feedback" condition). In order to further analyze the results, a two-way independent samples analysis of variance was conducted. Feedback and problem family were used as IV, (and the total number of correct answers for each individual was used as DV). Results showed that both feedback constraints and problem type had a statistically significant effect on the DV $F(2/115) = 3.6, p = .016$; $F(3/114) = 55.05, p < .001$. The interaction between the variables was non-significant $F(6/111) = .74, p = .621$. A post-hoc Games-Howell test did not detect statistical differences between feedback constraints (post-hoc test statistical error limit was established at .05). The same test detected statistical differences between the groups where the participants received the problems in the same order (direct or inverse) and the groups where model and target problem were presented in

different alphabetical order. A non-parametric post-hoc test was used because the data show a great amount variability both between and within

groups. The Games-Howell test is recommended under these circumstances (Hays, 1973).

Figure 1

Cumulative number of correct answers for each experimental group and condition

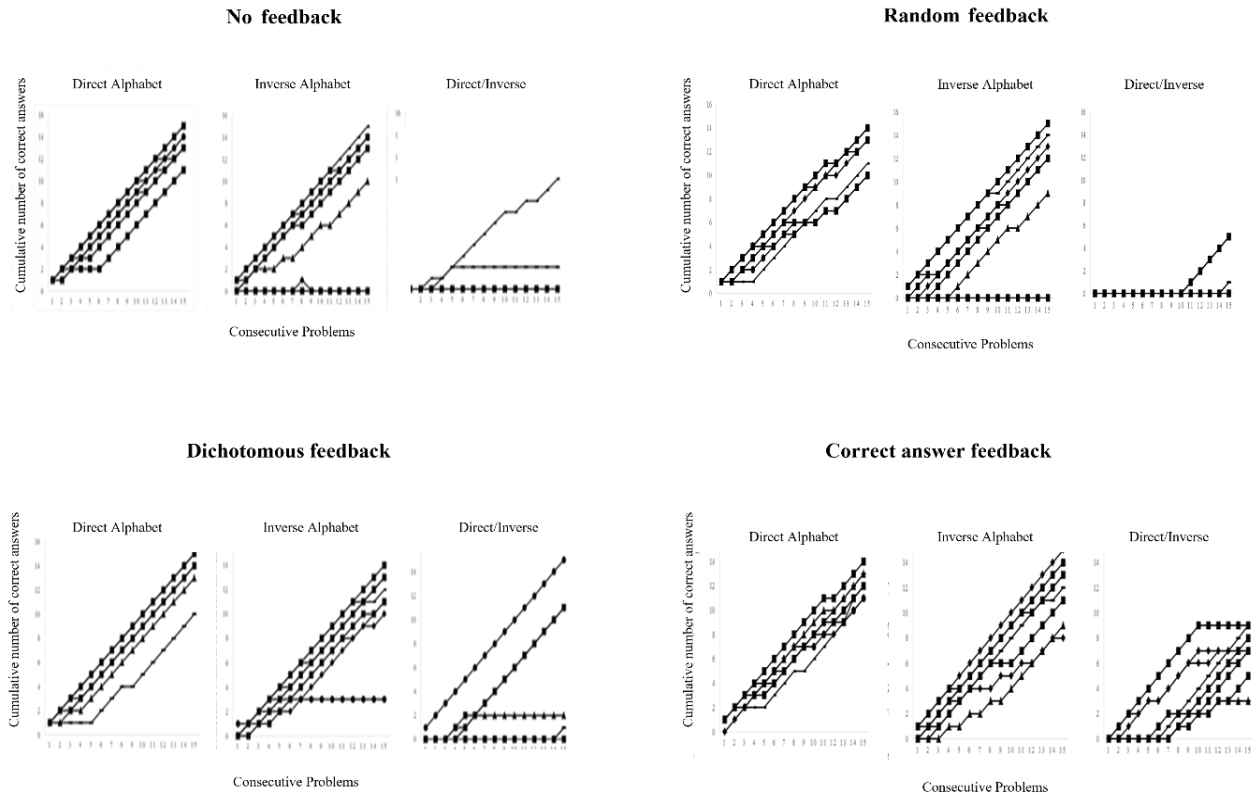


Table 1 shows the descriptive statistics for all experimental groups of the experiment. It presents these statistics, as most studies on FI use them as DV. Additionally, they provide a good synthesis for

the data presented on the first figure. For each condition, the table presents the statistics obtained in the first and last test problem.

Table 1
Descriptive Statistics, Experiment 1

Experimental Groups	ANCA		SDCA		%CA	
	First	Last	First	Last	First	Last
NF/D	1	13.5	0	1.6	100	90
NF/I	0.1	12.8	0.32	1.4	10	71.1
NF/DI	0.1	5.1	0.32	6.1	10	28.3
RF/D	1	12.3	0	1.8	100	82
RF/I	0.3	8.7	0.48	6.2	30	58
RF/DI	0	2.2	0	2.4	0	14.7
D/D	1	13.9	0	1.5	100	77.2
D/I	0.1	11.3	0.32	3.2	10	62.8
D/DI	0.1	5.2	0.32	6	10	28.9
CA/D	0.8	12.1	0.42	1	80	80.7
CA/I	0.4	12.2	0.52	2.3	40	77.3
CA/DI	0	6.5	0	2.2	0	42

Note. NF/D= No feedback/Direct; NF/I= No feedback/Invert; NF/DI= No feedback/Direct-Invert; RF/D= Random feedback/Direct; RF/I= Random feedback/invert; RF/DI= Random feedback/Direct-Invert; D/I= Dichotomus/Direct; D/I= Dichotomus/Invert; D/DI= Dichotomus/Direct-Invert; CA/D= Correct Answer/Direct; CA/I= Correct Answer/Invert; CA/DI= Correct Answer/Direct-Invert; ANCA= Average Number of Correct Answer; SDCA= Standard Deviation of Correct Answers; % CA= Percentage Correct Answers

Discussion

To discuss the findings in an orderly way, we will first center on the effects of the IV. After that, we will discuss the theoretical implications of those effects. The experiment assessed the effects of feedback and problem type on correct answer production. Apparently, feedback constraints have an effect on LSP solving, but the effect is hard to detect, (both visually and with the post-hoc test) because it only becomes apparent when the reader observes, exclusively, the Direct/Inverse conditions. These conditions confirm the experiment's first hypothesis; specifically lack of superficial similarities are easier to overcome by correct answer feedback. The number of correct responses then decreases systematically in the dichotomous feedback, no-feedback and random feedback conditions. The only discrepancy with the initial hypothesis was that it was expected that the dichotomous feedback condition be considerably more helpful than no-feedback (and the no-feedback condition, more similar to the random feedback condition). The results produced by the random feedback condition suggest that the accidental pairing of feedback/reinforcement with correct responses has little or no effect on their frequency. Regarding the second independent variable, the study attempted to proceed in an orderly manner by using as DV both problems where model and the target problem were in the same or different alphabetical order. Apparently, the first problems "may be too easy," and thus insensitive to FI. As mentioned previously Direct/Inverse LSP appear ideal for the assessment of FI (specifically if the research is focused on how to transcend superficial similarities). Although the experiment's hypothesis was confirmed by the data, the objective of determining the exact moment when structural solutions substitute superficial ones is hard to determine. All the groups where both model and target problems were presented in the same alphabetical order rarely presented incorrect responses (and thus it is difficult to draw conclusions from them). In a similar vein, as only the correct answer feedback condition presents a "more or less" equivalent combination of correct and incorrect answers, the generality of the finding may be questioned (also in three such conditions there is a conspicuous "floor" effect). In order to overcome these shortcomings, a second

experiment was conducted.

Regarding the similarities between the present and other studies, the data replicate the finding that more elaborate feedback may be more useful when implementing FI (Van der Kleij et al., 2015; Wisniewski et al., 2020). The results also coincide with Kluger and De Nisi's (1996) meta-analysis regarding the importance of task characteristics on FI. Finally, although operant conditioning may be an important component of FI (Carpenter & Vul, 2011; Grünke et al., 2017), and this learning mechanism is susceptible to non-contingent response-reinforcer pairings, the data showed no evidence of superstitious conditioning.

Experiment 2

The results of the first study suggest that Direct/Inverse LSP may be sensitive to FI; they also suggest that these interventions may help individuals transcend superficial similarities for structural ones. As only one out of four conditions presented this phenomenon clearly (correct answer feedback condition), the objective of the second experiment was to attempt a replication. Data from the first study also showed that the average number of correct responses is barely above 40%. One possible explanation for this result could be that 15 problems were simply insufficient to consolidate LSP solving by structural similarities. Thus, the second experiment attempted a replication of the correct answer feedback Direct/Inverse condition, using a higher number of problems. Another possible explanation for the relatively low number of correct responses obtained in the first experiment could be the lack of familiarity of the subjects with LSP (in the first study, participants received only one example of an LSP). Thus, in the second study, some participants were exposed to a training condition. In this condition, they received feedback on their performance solving different types of LSP. Additionally, the finding that dichotomous feedback had negligible effects on the Direct/Inverse condition is surprising, as other studies have found it useful in solving analogous LSP (Pulido, Hernández, et al., 2010). Thus, another objective of the second study was to assess the effects of dichotomous feedback on a greater number of LSP (and under conditions of previous training). Finally, most research on FI uses the number of correct answers as DV (Wisniewski et al., 2020).

This is understandable due to its mostly pragmatic focus. However, most basic research on learning phenomena uses both response and time indexes as DV's (Hull, 1943; Skinner, 1938; Spiliopoulos & Ortman, 2018; Staddon & Cerutti, 2003; Thorndike, 1898). Thus, in the second study, both correct answers and solution times were used as DV.

Methods

Participants

A sample of 69 undergraduate students participated in the study. The sample was predominantly female (62%), with an age mean of 20.2 years and a standard deviation of 1.2 years. Most students lived with their original families (93%). Inclusion criteria for the second study were the same as those used in experiment one.

Instruments

Participants were randomly assigned to one of four different experimental conditions. In all four conditions, the testing phase consisted of 20 letter-string problems, where the model was presented in direct alphabetical order (and the target in inverse alphabetical order). The testing problems were identical to the direct/inverse LSP presented in Appendix 1. The two control groups had no training phase. The two experimental groups had a training phase. During the training phase, participants received 16 different LSP of four distinct family types (the complete training problems are presented in Appendix 2). The first four training problems had both the model and the target problem in direct alphabetical order (ABC:ABD/MNO?). They were followed by four additional problems with both model and target problems in inverse alphabetical order (ONM:PNM/CBA?). A third group of four training problems consisted direct alphabetical order letter-strings, but the letters were non-consecutive (ACE/ACG:MOQ?). The training phase ended with four inverse order letter strings in non-consecutive order (QOM/ROM:ECA?). These letter-strings were piloted in a previous study (Pulido, Hernández et al., 2010) and their respective difficulty assessed. Their presentation order during the training phase was determined by difficulty

ratings (subjects first received the easiest problems, and gradually moved up to the most difficult ones).

In one experimental and one control condition, participants received dichotomous feedback in all problems (training or testing problems). In the remaining conditions, subjects received correct answer feedback. Feedback was delivered in the same way as described in the first experiment. As in the previous study, participants received all problems in printed booklets. After a revision of the applications, it was determined that nine experimental protocols were incorrectly applied in the correct answer experimental and control conditions. Thus, the data from the dichotomous feedback condition contains 39 subjects, while the correct answer feedback condition only 30 (15 in the experimental and 15 in the control groups).

Procedure

The experimental procedure was essentially the same as the one employed in the previous experiment. The only difference was that, as response times were used as DV, solution times were registered manually using a stopwatch. The stopwatch was initiated when the experimenter uncovered the problem; it was stopped as soon as the subject wrote down the answer.

Results

The left graph of Figure 2 shows the average response time for the dichotomous feedback conditions. The right graph shows the cumulative number of correct answers for these same conditions. Data from both the training and testing conditions are presented.

In general, the data in the left graph shows that solution times in the training and no-training conditions differ considerably. The former was relatively low, and the latter, relatively high. An independent samples *t* test confirms that the groups differed significantly $t(36) = 6.8, p < .001$. Regarding the number of correct answers, the training condition produces a considerably greater amount than the no training condition. An independent samples *t* test confirms that the conditions differed significantly $t(36) = 2.87, p = .007$.

Figure 2

Average response time and number of correct answers for each experimental

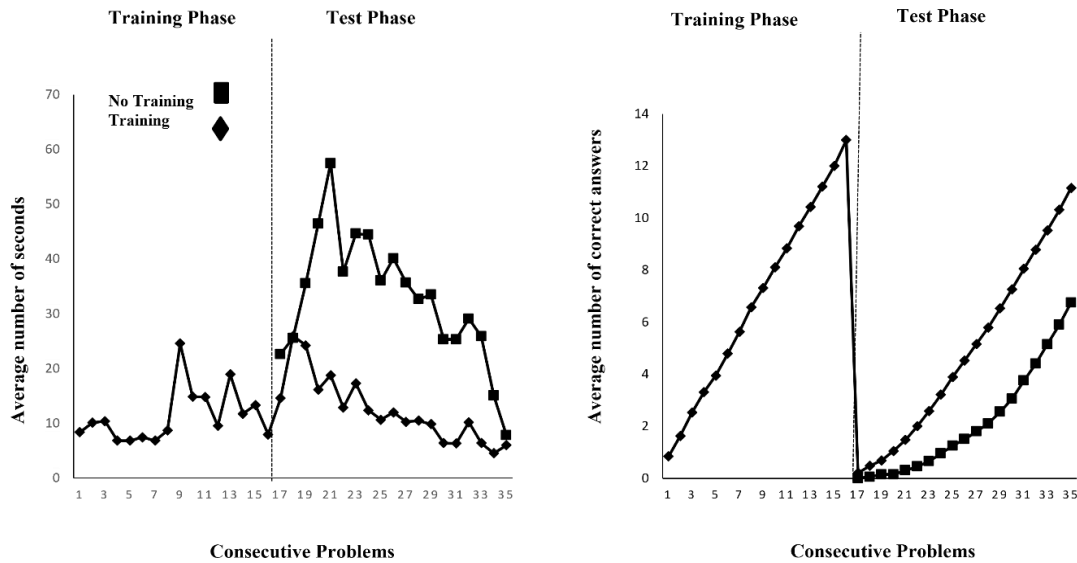


Table 2 shows the descriptive statistics for the test phase in the dichotomous feedback conditions

of experiment 2. Data from both the first and last problem are presented.

Table 2

Descriptive Statistics, Experiment 2. Dichotomous Feedback

Experimental Groups	ANCA		SDCA		%CA	
	First	Last	First	Last	First	Last
No Practice	0	6.8	0	1.4	0	34
Practice	0.1	11.2	0.42	3	20	56

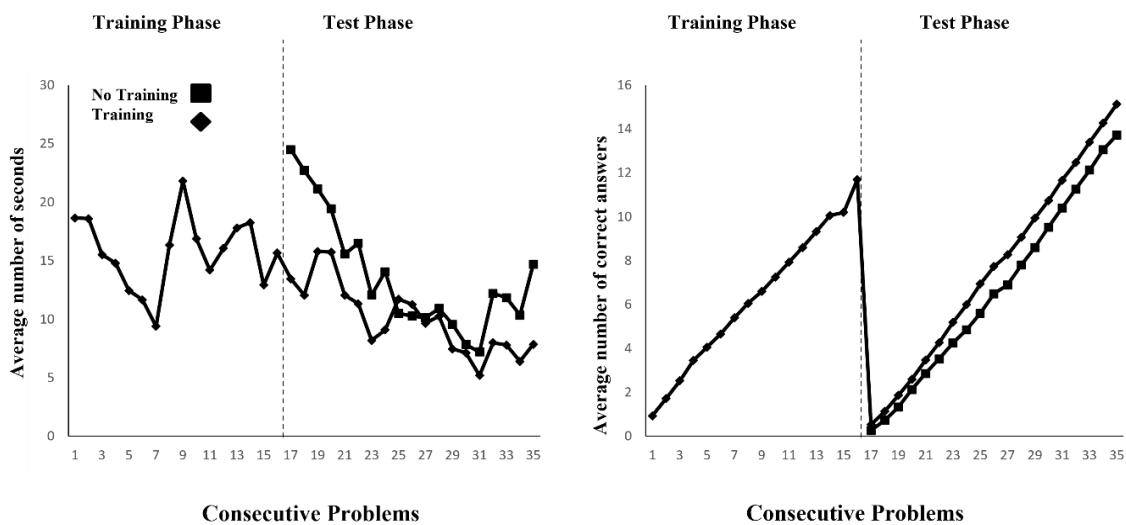
Note. ANCA= Average Number of Correct Answers; SDCA= Standard Deviation of Correct Answers; % CA= Percentage Correct Answers

The left graph of Figure 3 shows the average solution time for the correct answer feedback conditions; the right graph shows the cumulative

number of correct answers for these conditions. Data from both the training and testing conditions are shown.

Figure 3

Average response time and cumulative number of correct answers for each experimental condition



In general, Figure 3 shows that both average response times, and the average number of correct answers were similar for both the training and the no-training conditions. Additionally, both groups reached a high average number of correct answers (when compared with the results of the dichotomous feedback conditions). An independent samples *t* test showed that average response times on both the training and no training

conditions did not differ significantly $t(28) = 1.23, p = .23$; nor did the average number of correct responses $t(28) = .85, p = .402$.

Table 3 shows the descriptive statistics for the correct feedback conditions of the experiment. Data from the first and last problems are presented.

Table 3
Descriptive statistics, Experiment 2. Correct Answer Feedback

Experimental Groups	ANCA		SDCA		%CA	
	First	Last	First	Last	First	Last
No Practice	0.26	13.7	0.5	4.4	26.7	76.2
Practice	0.53	15.2	0.5	4.7	53.3	84.4

Note. ANCA = Average Number of Correct Answers; SDCA = Standard Deviation of Correct Answers; % CA = Percentage Correct Answers

Discussion

The results of the second study will be discussed in a similar way as that used in the first. Initially the main findings will be described; subsequently the theoretical implications of the findings will be discussed. Apparently, increasing the number of problems that the participants receive may also increase the average number of correct answers, (but only in the correct answer feedback conditions). In a similar vein, the data also confirm the finding of the first study, (in the sense that correct-answer feedback is associated with a higher number of correct answers than dichotomous feedback). Training has a strong beneficial and statistically significant effect in the dichotomous feedback conditions (although, as the participants’ progress through the problems, the initial advantage of the experimental over the control group decreases). The effects of training on the correct-answer feedback conditions are negligible.

The data suggest that there may be two approaches when teaching individuals analogous problems solving. A first approach involves exposing the individual to a large number of problems and providing scant information regarding performance. The data suggest that this approach is not practical, and that after more than 35 problems the average number of correct responses is only 56% (in the experimental group). Practice and richer feedback increase correct answer percentages in comparable conditions.

This finding is in general agreement with most of the scientific literature on feedback interventions (Deci et al., 1999; Hattie & Timperley, 2007; Kluger & DeNisi, 1996); and at odds with the literature that emphasizes practice (Christianson et al., 2012; Hegarty et al., 1995; Kim & Pak, 2002; Weaver & Kintsch, 1992).

The comparisons between experimental and control groups may help scientists value the effects of training on FI. Apparently, training may be useful when feedback provides poor information regarding task performance. The data also showed that training may be important during the early phases of task acquisition. Under a rich feedback context, or after “some experience” with the task, training appears to be an irrelevant manipulation. This finding helps to better understand training and to moderate the “overly enthusiastic” properties that have been attributed to the manipulation (Catrambone & Holyoak, 1989; Chen, 1999; Foreman-Murray & Fuchs, 2019; Gick & Holyoak, 1983; Novick & Holyoak, 1991).

It may be argued that the experiments conducted so far are “alien” to the FI scientific literature because each target problem receives some form of feedback, and this is infrequently the case in most FI studies. In the third experiment, this issue was addressed by assessing the effects of different training contingencies and subsequently observing LSP solving without any form of feedback.

Experiment 3

Since the early research on FI, scientists have argued that the ultimate goal should be to produce “independent problem solvers” (Ashford & Cummings, 1983; Balcazar et al., 1985). This means that the subject must reach the correct answer “unprompted” (Grose & Birney, 1963; Reed et al., 1974). Thus, it is arguable that the studies conducted so far belong within the domain of the FI literature. In order to reconcile the present study with the FI literature, in experiment 3, individuals received training for a variable number of problems. Subjects were subsequently asked to solve 9 problems on their own. Experiment 3 was also conducted to further assess an idea that started to develop itself in the second study. Specifically, the data suggested that given the correct answer solution to a problem, individuals quickly began to produce structural solutions when solving letter-strings. In sum, given rich feedback, the beneficial effects of practice quickly become asymptotical. In order to assess this hypothesis, the third experiment exposed participants to one, five, or ten practice problems with correct-answer feedback. Subsequently they began a test phase where no feedback was received.

Methods

Participants

A total of 30 undergraduate students participated in the study. The sample was predominantly masculine (66.6%), and the average age was determined at 23.1 years with a

SD of 2.1 years. Half of the students lived with their families (50%). Inclusion criteria for the third study were the same as those used in experiment one.

Instruments

Participants were randomly assigned to one of three different experimental conditions. In the first, individuals received one model and one target problem. Once the participant produced and answer, he/she was given feedback regarding his performance (correct or incorrect); if incorrect, the subject was shown the correct answer, by means of a white paper card. In the second condition, participants received feedback on five different consecutive problems. In the third condition, participants received feedback on ten such problems. In all training problems, the model letter string was presented in direct alphabetical order (the target string was presented in inverse order). The testing phase consisted of nine new letter string problems, where the model was in direct alphabetical order, and the target was in inverse alphabetical order (both training and test problems were identical to the direct/inverse LSP presented in Appendix 1).

Procedure

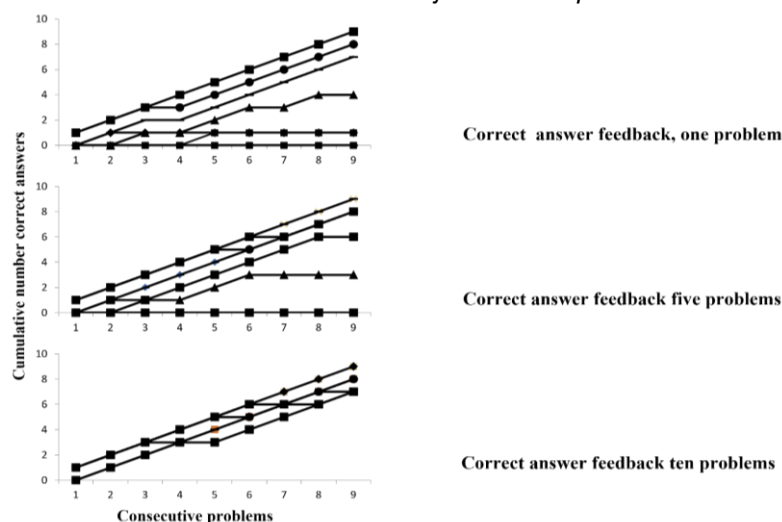
The procedure was the same used in the previous studies.

Results

Figure 4 shows the cumulative number of correct answers for all subjects and conditions.

Figure 4

Cumulative number of correct answers for each subjects and experimental condition. Test phase



In general, the data in Figure 4 show that the cumulative number of correct answers during the test phase is a direct function of the number of training problems; performance variability is an inverse function of training problems. A one-way independent samples ANOVA test showed differences between the groups $F(2/27) = 5.25, p = .01$. A Games-Howell post-hoc test showed that

the five and ten practice problems groups significantly differed from the problem one practice group ($p < .05$).

Table 4 shows the descriptive statistics for experiment 3. Data from the first and last testing problems are presented.

Table 4
Descriptive Statistics Experiment 3 Test Problems

Experimental Groups	ANCA		SDCA		%CA	
	First	Last	First	Last	First	Last
One training problem	0.3	3.9	0.5	4	30	43
Five training problems	0.4	6.6	0.5	3.1	40	72.8
Ten training problems	0.7	8	0.5	0.8	70	88.9

Note. ANCA = Average Number of Correct Answers; SDCA = Standard Deviation of Correct Answers; % CA = Percentage Correct Answers

Discussion

In general, the data suggest that the overall percentage of correct answers is a direct function of the number of training problems. The group percentages of correct answers show an important improvement when moving from one to five training problems (29.8%). The improvement becomes more modest when moving from five to ten problems (16.1%). The percentage of correct answers obtained for the five and ten training problem conditions are very similar to those obtained for the correct answer conditions in the second study.

In sum, the results from the third study suggest that, even in the absence of feedback during the testing phase, training problems with correct-answer feedback may help individuals solve problems, based on structural features. This procedure shows results that are very similar to those obtained in the second study, thus suggesting that the most important “learning moments” occur during the early exposure to the problems (during the first 10 problems). The findings help the present study reconcile with the more traditional procedures used in the FI literature (Wisniewski et al., 2020). The results are also similar to most of the psychological literature on the learning process. This scientific literature suggests that the most important gains in response acquisition occur during the early exposure of the organism to the experimental

contingencies (Rescorla & Wagner, 1972; Thorndike, 1898).

General Discussion

The present study addressed a number of issues; thus, the first objective of this discussion will be to describe the principal results of the three studies in an orderly fashion. The first study may be conceptualized as a pilot study that allowed three basic things. First, it permitted the authors to confirm previous findings that suggest that LSP differ in difficulty, and that some of them may require a structural understanding of the problem (Pulido, Almaraz, et al., 2010; Pulido, Hernández et al., 2010). In second place, the first study showed that LSP respond to FI in a similar way as other type of problems. Specifically, as FI become more informative their “beneficial” effects on problem solving also increase (Hattie & Zierer, 2019; Kluger & DeNisi, 1996; Wisniewski et al., 2020). Finally, the first study allowed the authors to evaluate their specific repeated measures experimental design and assess control group conditions that had not been used in previous studies. Regarding the former, results show that the experimental design could allow scientists to identify crucial events affecting the learning curve given that LSP are hard, and a floor effect does not “hide” them. Regarding the later, FI research has shown that operant conditioning is an important component of the intervention (Carpenter & Vul, 2011; Grünke et al., 2017; King

et al., 2000), thus, a lack of superstitious conditioning using the non-contingent control group is a curious finding; perhaps future studies may provide an explanation.

The second study was basically conceived by building upon the lessons learned from the first one. As only direct/inverse problems seemed to benefit from FI, only those problems were used. Additionally, as the first study showed only a relatively modest percentage of correct answers (42% in the correct answer condition), the second assessed if longer exposure to the FI and including a training condition would improve upon this figure. The second study also tried to have a second look at the dichotomous feedback condition under these new experimental conditions. The first study also left the authors "wanting" for finer look at the precise moment when the FI had its most dramatic effects on LSP solving. For this reason, solution time was added as a DV, and group averages (rather than individual performance) were presented in the figures. Results showed that these new conditions could increase the number of correct answers (84.4% in the correct answers feedback with training condition). The new experimental conditions also made important improvements in the dichotomous feedback condition. The comparison of the two feedback conditions also produced an interesting pragmatic result. Specifically, previous training helps problem solving only in the absence of informative feedback. Finally, the second study actually allowed a closer look at the effects of FI on LSP solving but only in the dichotomous feedback condition (correct answers and solution times in the other condition simply proceeded in very straightforward and lineal fashion). Data from dichotomous FI suggest that most learning occurs within the first ten problems. After that, both the average number of correct answers and the average solution times, in both trained and untrained individuals, increases asymptotically (average solution times actually "touch" at the final problem).

Finally, the third study was developed with the objective of connecting the findings with more traditional FI studies. Specifically assessing the manipulations identified in the present study on unprompted LSP. This study further assessed the finding of the second study regarding the specific moment when FI affects LSP solving. The study

showed that correct answer feedback has beneficial effects even on problems where no feedback is provided. The experiment suggested, as did the second one, that the benefits of FI occur during the first problems and subsequently may become asymptotic.

Therefore, what do these studies suggest about how FI help individuals substitute problem solving based on superficial similarities for structural ones? The answer to these questions is based on a number of assumptions, and the present authors are obliged to present them clearly to the readers. First, a basic premise of the experiments is that direct/inverse LSP require a structural understanding of the problem setting, in order to be "properly" solved. This assumption is based on another one. Specifically, that difficulty indexes presented in the first study (and in other previously published ones, Pulido, Almaraz, et al., 2010; Pulido, Hernández, et al., 2010) are an indication that superficial similarities are simply insufficient to solve. A qualitative study of the discourse produced by individuals attempting to solve this type of problems (and other LSP), also supported this assumption (Pulido, 2002). Some readers may find these arguments compelling (others may not). However, should these premises be accepted, a number of theoretical and pragmatic issues derive from the experiments. First, the non-contingent control group of the first experiment showed the lowest percentage of correct answers. As this procedure is helpful in the development of behavior repertoires in basic operant conditioning research (Brown & Jenkins, 1968; Meador & Hatfield, 2010; Timberlake & Grant, 1975); the role of operant conditioning in FI may be less important than has been previously suggested (Carpenter & Vul, 2011; Grünke et al., 2017; King et al., 2000). A second conclusion is relevant to the current dispute regarding the importance of practice on problem solving. There is currently a "hotly" contested debate regarding how problem solving should be taught to individuals in schools, digital platforms, and other learning contexts. This debate can be summarized as a quantity vs quality dispute (Ball et al., 2010; Christianson et al., 2012; Doabler et al., 2019; Kim & Pak, 2002; Lei, 2010). The data produced by these studies favor the former alternative. Specifically, it suggests that individuals may acquire the ability to produce structural solutions to problems given informative feedback. However,

even in the absence of “very informative” feedback (dichotomous feedback), structural solutions may be produced with limited training. The finding is not irrelevant, given that even a brief look at how “core” problem solving skills are taught in different countries shows great heterogeneity in teaching policies, and that many of them favor quantity over quality (Bolton, 2019; European Education and Culture Executive Agency et al., 2012; U.S. Department of Education, 2003). Finally, although repeated measures designs became unpopular after the “demise” of behaviorism, the results of the present studies suggest that their power to evaluate the effects of IV over DV in education is interesting and may help scientists make better decisions within the behavioral sciences.

Lastly, this study must recognize a number of limitations. First, only letter string problems were used. Future studies may assess the generality of the findings. In second place, both problem presentation and DV measurement could probably be more precise if the problems were presented in a digital platform.

Acknowledgements

The authors wish to thank the support of Ricardo Rivas, Perla Vázquez, Ernesto Reyes, Alejandro González and Diana Betancourt.

References

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, 48(2), 87-113. <http://dx.doi.org/10.1080/00461520.2013.775712>
- Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, 41(5), 573-583. <https://doi.org/10.1016/j.ergon.2011.05.001>
- Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32(3), 370-398. [https://doi.org/10.1016/0030-5073\(83\)90156-3](https://doi.org/10.1016/0030-5073(83)90156-3)
- Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3-4), 65-89. https://doi.org/10.1300/J075v07n03_05
- Ball, L. J., Hoyle, A. M., & Towse, A. S. (2010). The facilitatory effect of negative feedback on the emergence of analogical reasoning abilities. *British Journal of Developmental Psychology*, 28(3), 583-602. <https://doi.org/10.1348/026151009X461744>
- Bokhove, C., & Drijvers, P. (2012). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education*, 58(1), 197-208. <https://doi.org/10.1016/j.compedu.2011.08.010>
- Bolton, L. (2019). Foundational mathematics education in developing countries. K4D Helpdesk Report 657. Brighton, UK: Institute of Development Studies. <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14736>
- Brown, P. L., & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior*, 11(1), 1-8. <https://doi.org/10.1901/jeab.1968.11-1>
- Burns, B. D. (1996). Meta-analogical transfer: Transfer between episodes of analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(4), 1032-1048. <https://doi.org/10.1037/0278-7393.22.4.1032>
- Carpenter, S. K., & Vul, E. (2011). Delaying feedback by three seconds benefits retention of face-name pairs: The role of active anticipatory processing. *Memory & Cognition*, 39(7), 1211-1221. <https://doi.org/10.3758/s13421-011-0092-1>
- Catania, A. C. (1998). *Learning* (4th ed.). Upper Saddle River, Prentice Hall.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1147-1156. <https://doi.org/10.1037/0278-7393.15.6.1147>
- Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology*, 91(4), 703-715. <https://doi.org/10.1037/0022-0663.91.4.703>
- Chi, M. T., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, 47(3), 177-188. <https://doi.org/10.1080/00461520.2012.695709>
- Christianson, K., Mestre, J. P., & Luke, S. G. (2012). Practice makes (nearly) perfect: Solving 'students-and-professors'-type algebra word problems. *Applied Cognitive Psychology*, 26(5), 810-822. <https://doi.org/10.1002/acp.2863>
- Corbalan, G., Paas, F., & Cuyppers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*, 55(2), 692-703. <https://doi.org/10.1016/j.compedu.2010.03.002>
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481. <https://doi.org/10.3102/00346543058004438>
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation.

- Psychological Bulletin*, 125(6), 692-700.
<https://doi.org/10.1037/0033-2909.125.6.627>
- Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 198-215.
<https://doi.org/10.1006/obhd.1995.1043>
- Doabler, C. T., Clarke, B., Kosty, D., Turtura, J. E., Firestone, A. R., Smolkowski, K., Jungjohann, K., Brafford, T. L., Nelson, N. J., Sutherland, M., Fien, H., & Maddox, S. A. (2019). Efficacy of a First-Grade Mathematics Intervention on Measurement and Data Analysis. *Exceptional Children*, 86(1), 77–94. <https://doi.org/10.1177/0014402919857993>
- European Education and Culture Executive Agency, Eurydice, Forsthuber, B., Horvath, A., Almeida Coutinho, A. (2012). *Science education in Europe – National policies, practices and research*, Publications Office. <https://data.europa.eu/doi/10.2797/7170>
- Foreman-Murray, L., & Fuchs, L. S. (2019). Quality of explanation as an indicator of fraction magnitude understanding. *Journal of Learning Disabilities*, 52(2), 181-191.
<https://doi.org/10.1177/0022219418775120>
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1-38. [https://doi.org/10.1016/0010-0285\(83\)90002-6](https://doi.org/10.1016/0010-0285(83)90002-6)
- Goldwater, M. B., & Jamrozik, A. (2019). Can a relational mindset boost analogical retrieval? *Cognitive Research: Principles and Implications*, 4(47), 1-16. <https://doi.org/10.1186/s41235-019-0198-8>
- Grose, R. F., & Birney, R. C. (1963). *Transfer of learning: An enduring problem in psychology*. Van Nostrand.
- Grünke, M., Sperling, M., & Burke, M. D. (2017). The Impact of Explicit Timing, Immediate Feedback, and Positive Reinforcement on the Writing Outcomes of Academically and Behaviorally Struggling Fifth-Grade Students. *Insights into Learning Disabilities*, 14(2), 135-153.
<https://files.eric.ed.gov/fulltext/EJ1164944.pdf>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hattie, J., & Zierer, K. (2019). *Visible learning insights*. Routledge.
- Hays, W. L. (1973). *Statistics for the social sciences*. Holt, Rinehart & Winston.
- Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, 87(1), 18-32.
<https://doi.org/10.1037/0022-0663.87.1.18>
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton- Century Crofts.
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2020). Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement. *Higher Education*, 137-156.
<https://doi.org/10.1007/s10734-019-00469-2>
- Iversen, I. H., & Lattal, K. A. (1991). *Experimental analysis of behavior*. Elsevier.
- Johnson, D. A. (2013). A component analysis of the impact of evaluative and objective feedback on performance. *Journal of Organizational Behavior Management*, 33(2), 89-103.
<https://doi.org/10.1080/01608061.2013.785879>
- Jonassen, D. H. (1997). Instructional design models for well-structured and Ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45, 65-94.
<https://doi.org/10.1007/BF02299613>
- Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology: A systematic text in the science of behavior*. Appleton-Century Crofts
- Kelley, M. E., Nadler, C. B., Rey, C., Cowie, S., & Podlesnik, C. A. (2017). Noncontingent reinforcement competes with response performance. *Journal of the Experimental Analysis of Behavior*, 107(3), 343-353.
<https://doi.org/10.1002/jeab.255>
- Kim, E., & Pak, S.-J. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics*, 70(7), 759-765. <https://doi.org/10.1119/1.1484151>
- King, P. E., Young, M. J., & Behnke, R. R. (2000). Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education*, 49(4), 365-374.
<https://doi.org/10.1080/03634520009379224>
- Kirk R. (1995). *Experimental design: Procedures for the behavioral sciences*. 3rd. Pacific Grove, Brooks/Cole.
- Kluger, A. N., & 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.
<https://doi.org/10.1037/0033-2909.119.2.254>
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279-308.
<https://doi.org/10.1007/BF01320096>
- Lei, J. (2010). Quantity versus quality: A new approach to examine the relationship between technology use and student outcomes. *British Journal of Educational Technology*, 41(3), 455-472.

- <https://doi.org/10.1111/j.1467-8535.2009.00961.x>
 Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences*, 15(4), 431-449.
https://doi.org/10.1207/s15327809jls1504_1
- Mangiapanello, K. A., & Hemmes, N. S. (2015). An analysis of feedback from a behavior analytic perspective. *The Behavior Analyst*, 38(1), 51-75.
<https://doi.org/10.1007/s40614-014-0026-x>
- Meador, S. K., & Hatfield, D. B. (2010). *The Corsini Encyclopedia of Psychology*. Wiley Online Library.
<https://doi.org/10.1002/9780470479216.corpsy0961>
- Mohamed, A. (2020). Feedback in Computer-Assisted Language Learning: A Meta-Analysis. *The Electronic Journal fo English as a Second Language*, 24(2). <http://www.tesl-ej.org/wordpress/issues/volume24/ej94/ej94a4/>
- Murphy, P. (2010). Web-based collaborative reading exercises for learners in remote locations: the effects of computer-mediated feedback and interaction via computer-mediated communication. *ReCALL*, 22(2), 112-134.
<https://doi.org/10.1017/S0958344010000030>
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3), 398-415.
<https://doi.org/10.1037/0278-7393.17.3.398>
- Patchan, M. M., & Puranik, C. S. (2016). Using tablet computers to teach preschool children to write letters: Exploring the impact of extrinsic and intrinsic feedback. *Computers & Education*, 102, 128-137.
<https://doi.org/10.1016/j.compedu.2016.07.007>
- Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst*, 5(1), 101-102.
<https://doi.org/10.1007/BF03393144>
- Pulido, M. A. (2002). Transferencia analógica en la solución de problemas de cuatro términos con un diseño de ensayo múltiple. *Revista Mexicana de Psicología*, 19(1), 23-34.
- Pulido, M. A., Almaraz, D. A., García, D. G., & Martínez, L. M. (2010). Ten years of research on letter string problem solving by Meta-Analogical Transfer. *Journal of Behavior, Health & Social Issues*, 2(1), 83-89. <https://doi.org/10.5460/jbhsi.v2.i1.09>
- Pulido, M. A., de la Garma Valenzuela, M., & Vargas, C. P. (2010). Solución de problemas por transferencia analógica: efectos de la disponibilidad del modelo. *Revista Intercontinental de Psicología y Educación*, 12(1), 157-170.
<https://www.redalyc.org/articulo.oa?id=80212393009>
- Pulido, M. A., Hernández, C., Peña, L. M., Ponce, N. M., & Rebolledo, M. (2010). El condicionamiento operante de la solución de problemas de secuencias de letras. *Revista Mexicana de Análisis de la Conducta*, 36(3), 55-76.
<https://doi.org/10.5514/rmac.v36.i3.04>
- Pulido, M. A., Rosas, É. O., & Piñol, N. L. (2005). La solución de problemas de cuatro términos por transferencia analógica: El efecto de las similitudes superficiales y del tipo de problema. *Revista Mexicana de Psicología*, 22(2), 433-440.
<https://www.redalyc.org/articulo.oa?id=243020634007>
- Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, 6(3), 436-450.
[https://doi.org/10.1016/0010-0285\(74\)90020-6](https://doi.org/10.1016/0010-0285(74)90020-6)
- Reimers, F., & Chung, C. K. (2016). *Enseñanza y aprendizaje en el siglo XXI: metas, políticas educativas y currículo en seis países*. Fondo de Cultura Económica.
- Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). Appleton-Century-Crofts.
- Riper, H., van Straten, A., Keuken, M., Smit, F., Schippers, G., & Cuijpers, P. (2009). Curbing problem drinking with personalized-feedback interventions: a meta-analysis. *American Journal of Preventive Medicine*, 36(3), 247-255.
<https://doi.org/10.1016/j.amepre.2008.10.016>
- Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 427-440.
<https://doi.org/10.1037/0278-7393.23.2.427>
- Ryan, T. P. (2013). *Sample size determination and power*. John Wiley & Sons.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.
<https://doi.org/10.1007/BF00117714>
- Schwartz, D. L., Chase, C. C., Opezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759-775.
<https://doi.org/10.1037/a0025140>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
<https://doi.org/10.3102/0034654307313795>
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Basic Books.
- Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century.

- Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, 38(2), 168-172. <https://doi.org/10.1037/h0055873>
- Snyder, L. G., & Snyder, M. J. (2008). Teaching critical thinking and problem solving skills. *The Journal of Research in Business Education*, 50(2), 90-99. http://dme.childrenshospital.org/wp-content/uploads/2019/02/Optional_Teaching-Critical-Thinking-and-Problem-Solving-Skills.pdf
- Spiliopoulos, L., & Ortmann, A. (2018). The BCD of response time analysis in experimental economics. *Experimental Economics*, 21, 383–433. <https://doi.org/10.1007/s10683-017-9528-1>
- Staddon, J. E., & Cerutti, D. T. (2003). Operant conditioning. *Annual Review of Psychology*, 54, 115-144. <https://doi.org/10.1146/annurev.psych.54.1016.01.145124>
- Thorndike, E. L. (1898). Animal intelligence: an experimental study of the associative processes in animals. *Psychological Review Monograph Supplements*, 2(4), i-109. <https://doi.org/10.1037/h0092987>
- Thorndike, E. L. (1903). *Educational psychology*. Lemcke & Buechner. <https://doi.org/10.1037/10528-000>
- Timberlake, W., & Grant, D. L. (1975). Auto-shaping in rats to the presentation of another rat predicting food. *Science*, 190(4215), 690-692. <https://doi.org/10.1126/science.190.4215.690>
- Toader, E.-A., & Lungu, M.-A. (2015). The impact of elaborated feedback on competency assessment of IT professionals. *Informatica Economica*, 19(3), 80-88. <http://www.revistaie.ase.ro/content/75/07%20-%20Toader,%20Lungu.pdf>
- U.S. Department of Education, National Center for Education Statistics. The Condition of Education 2003, NCES 2003–067. <https://nces.ed.gov/pubs2003/2003067.pdf>
- Van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272. <https://doi.org/10.1016/j.compedu.2011.07.020>
- Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475-511. <https://doi.org/10.3102/0034654314564881>
- Van der Kleij, F. M., Timmers, C. F., & Eggen, T. J. (2011). The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review. *Cadmo*, 21-38. <https://doi.org/10.3280/CAD2011-001004>
- Weatherly, N. L., & Malott, R. W. (2008). An analysis of organizational behavior management research in terms of the three-contingency model of performance management. *Journal of Organizational Behavior Management*, 28(4), 260-285. <https://doi.org/10.1080/01608060802454643>
- Weaver, C. A., & Kintsch, W. (1992). Enhancing students' comprehension of the conceptual structure of algebra word problems. *Journal of Educational Psychology*, 84(4), 419-428. <https://doi.org/10.1037/0022-0663.84.4.419>
- Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International Encyclopedia of Education* (2nd ed., pp.5738-5745). Pergamon.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Zeiler, M. D. (1972). Superstitious behavior in children: An experimental analysis. *Advances in Child Development and Behavior*, 7, 1-29. [https://doi.org/10.1016/S0065-2407\(08\)60438-3](https://doi.org/10.1016/S0065-2407(08)60438-3)

Appendix 1

Problems used in Experiment 1

Direct/Direct	Inverse/Inverse	Direct/Inverse
ABC:ABD/LMN?	ONM:PNM/NML?	ABC:ABD/NML?
ABC:ABD/RST?	ONM:PNM/TSR?	ABC:ABD/TSR?
ABC:ABD/OPQ?	ONM:PNM/QPO?	ABC:ABD/QPO?
ABC:ABD/TUV?	ONM:PNM/VUT?	ABC:ABD/VUT?
ABC:ABD/MNO?	ONM:PNM/ONM?	ABC:ABD/ONM?
ABC:ABD/DEF?	ONM:PNM/FED?	ABC:ABD/FED?
ABC:ABD/FGH?	ONM:PNM/HGF?	ABC:ABD/HGF?
ABC:ABD/UVW?	ONM:PNM/WXU?	ABC:ABD/WXU?
ABC:ABD/HIJ?	ONM:PNM/JIH?	ABC:ABD/JIH?
ABC:ABD/PQR?	ONM:PNM/RQP?	ABC:ABD/RQP?
ABC:ABD/STU?	ONM:PNM/UTS?	ABC:ABD/UTS?
ABC:ABD/XYZ?	ONM:PNM/ZYX?	ABC:ABD/ZYX?
ABC:ABD/NOP?	ONM:PNM/PON?	ABC:ABD/PON?
ABC:ABD/JKL?	ONM:PNM/LKJ?	ABC:ABD/LKJ?
ABC:ABD/CDE?	ONM:PNM/EDC?	ABC:ABD/EDC?

Appendix 2

Training Problems used in Experiment 2

Direct/Direct Consecutive	Inverse/Inverse Consecutive	Direct/Direct Non-Consecutive	Inverse/Inverse Non Consecutive
First Training	Second Training	Third training	Fourth Training
ABC:ABD/LMN?	ONM:PNM/NML?	ACE:ACG/MOQ?	QOM:ROM/ECA?
ABC:ABD/RST?	ONM:PNM/TSR?	ACE:ACG/HJL?	QOM:ROM/LJH?
ABC:ABD/OPQ?	ONM:PNM/QPO?	ACE:ACG/RTV?	QOM:ROM/VTR?
ABC:ABD/TUV?	ONM:PNM/VUT?	ACE:ACG/DFH?	QOM:ROM/HDF?