

DTIC FILE COPY

②

AD-A222 324

FELICITY CONDITIONS FOR COGNITIVE
SKILL ACQUISITION: TUTORIAL
INSTRUCTION DOES NOT NEED THEM

Technical Report PCG - 13

DEPARTMENT
of
PSYCHOLOGY



DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Carnegie-Mellon University

DTIC
ELECTE
JUN 06 1990
S E D
Co

90 06 05 061

**FÉLICITY CONDITIONS FOR COGNITIVE
SKILL ACQUISITION: TUTORIAL
INSTRUCTION DOES NOT NEED THEM**

Technical Report PCG - 13

Kurt VanLehn
Departments of Psychology & Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 U.S.A.

February 19, 1988
(revised: April 7, 1989)

This is the final report on the research supported by the Personnel and Training Research Program, Psychology Sciences Division, Office of Naval Research, under contract N00014-86-K-0349. Reproduction in whole or in part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited. The author is grateful for the help of Micki Chi, whose close reading of early drafts were invaluable.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited		
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
4. PERFORMING ORGANIZATION REPORT NUMBER(S) PCG - 13			7a. NAME OF MONITORING ORGANIZATION Personnel & Training Research, Psychological Sciences Div., Offical of Naval Research		
6a. NAME OF PERFORMING ORGANIZATION CarnegieMellon University		6b. OFFICE SYMBOL (if applicable)		7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000	
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213		8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as Monitoring Orgnization		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0349	
8b. OFFICE SYMBOL (if applicable)		10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO N/A		PROJECT NO. N/A	TASK NO. N/a
8c. ADDRESS (City, State, and ZIP Code)		WORK UNIT ACCESSION NO N/a			
11. TITLE (Include Security Classification) Felicity conditions for cognitive skill acquisition: Tutorial instruction does not need them. Unclassified					
12. PERSONAL AUTHOR(S) Kurt VanLehn					
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM 86 June 15 TO 87 Oct 31		14. DATE OF REPORT (Year, Month, Day) 1988 February 19	
15. PAGE COUNT 20					
15. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Skill acquisition, training, curriculum design, scope and sequence. (C) 1		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) SEE REVERSE					
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION		
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Susan Chipman			22b. TELEPHONE (Include Area Code) (202) 696-4322		22c. OFFICE SYMBOL

Abstract

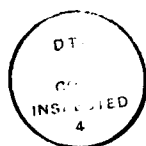
Theoretical work suggests that when students learn a complex skill, they may face ambiguities in how to interpret the training material, and that there may be social conventions, called felicity conditions, about how the teacher will provide information that help the students resolve these ambiguities. One proposed felicity condition is for the teacher to guarantee that a separate lesson will be used for the introduction of new methods or concepts that are disjunctively related to the previously taught material. This hypothesized felicity condition, called one-disjunct-per-lesson, is tested in two experiments. Fourth-grade students were taught multidigit multiplication in two conditions, one that obeyed the felicity condition and one that violated it. It was expected that the training condition that violated the felicity condition would cause greater confusion, but this did not occur. Surprisingly, students in that condition did better than students in the one-disjunct-per-lesson condition on a transfer task. Some revisions to the felicity condition view are suggested in order explain this unexpected result.

Knowledge

(A)

A: 12/12

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist _____ and/or	
Special	
A-1	



1. Introduction

While constructing a model of arithmetic skill acquisition (VanLehn, 19??, VanLehn, 1983), I noticed that arithmetic textbooks uniformly obey a convention that I call *one-disjunct-per-lesson*. If the procedure to be taught has two or more subprocedures that are disjunctively related (either one or the other is used), then the textbooks introduce at most one of these subprocedure per lesson. For instance, the procedure for subtracting multidigit whole numbers has three disjunctively related subprocedure for processing a column:

1. If the column has just one digit, write it in the answer.
2. If the column's top digit is smaller than the bottom digit, then borrow.
3. Otherwise, take the difference between the columns two digits and write it in the answer.

Suppose that students have mastered the third subprocedure, but have not yet been introduced to the first two. They can correctly solve $56-23$ but not $56-3$, whose solution utilizes the first subprocedure, nor $56-29$, whose solution utilizes the second subprocedure, nor $56-9$, whose solution utilizes both subprocedures. One could write a textbook that introduces both subprocedures 1 and 2 in the same lesson. It could, for instance, use the three problems just mentioned. Such a textbook would violate the *one-disjunct-per-lesson* convention. None of the textbooks I examined violated the *one-disjunct-per-lesson* convention. I examined the textbooks from Heath and Scotts-Foresman, because these texts were used by the schools involved in the study, and from three other major publishers. In only one case was there any doubt about conformance with the convention. In the third grade book of the 1975 Scotts-Foresman series, there is a two page lesson that introduces subprocedure 1 on one page and subprocedure 2 on the other. Whether this actually violates the *one-disjunct-per-lesson* convention depends on how "lesson" is defined, an issue that is discussed at length later. However, except for this one case, all of the other cases of subprocedure introduction took place in a lesson that introduced just one subprocedure.

Why would there be such conformance to the *one-disjunct-per-lesson* convention? Arithmetic textbooks have evolved over the past two centuries under the influence of many theories, experimental results and practical experiences. Although the convention could simply be a fad among textbook designers, it is more likely that textbooks written in conformance with the

convention facilitate learning. This would explain its ubiquity.

Why would teaching at most one disjunct per lesson facilitate learning? Several possible explanations come to mind. Teaching two subprocedures makes the explanations more complex, and presumably more difficult to understand. The text (and teacher) would have to say, "Today we're going to learn two things. The first is X, which you should use only when A. The second is Y, which you should use only when B." Further descriptions of the two subprocedures should always be prefaced by a remark indicating which subprocedure is being described. Although plausible, this cannot be the only explanation of the one-disjunct-per-lesson convention, for there is growing evidence that students attend more to the worked example exercises than to linguistic descriptions of problem solving procedures (Badre, 1972; LeFevre & Dixon, 1986; Anderson, Farrell, & Saurers, 1984; VanLehn, 1986). Anderson's group, for example, has found that students tend to answer exercise problems by finding a similar exercise in the book that has been solved already, then mapping the solution over to their problem (Anderson, Farrell, & Saurers, 1984; Pirolli & Anderson, 1985). Similar phenomena have been found by other investigators (Chi, Bassok, Lewis, Reimann & Glaser, 19??). If this learning process is the one in use for arithmetic, then a curriculum that puts two or more subprocedures in a lesson is going to make the students' task harder. They will have two types of solved exercises in the lesson, and they will have to decide which to use as the source for their analogical problem solving. If they decide incorrectly, they could develop serious misconceptions. On the other hand, if all lessons have just one subprocedure in them, then the students do not have to make a choice. They simply refer to any of the solved examples in the lesson. In short, regardless of whether students attend more to the examples or the linguistic descriptions, it appears that introducing more than one disjunct in the same lesson harms the ability of the lesson to communicate the new material to the student.

If we take seriously the idea that learning involves communication of information, then instruction should have conventions that govern it, since all other forms of communications between people seem to. The fact that the students and teachers do not have conscious access to the rules governing the communication between them is not an argument that such rules do not exist. In fact, it is circumstantial evidence that such rules *do* exist, because most rules of human communication are not available to conscious access. In honor of some famous tacit conventions on natural language conversation, Austin's (1962) *felicity conditions*, the conjecture that there might

be tacit conventions governing instruction is called the *felicity conditions conjecture*.

There are some formal results in learning theory that indicate the value of felicity conditions and the one disjunct per lesson constraint in particular. Currently, Valiant's (1984) criterion is held to be an excellent definition of what it means to learn a concept from examples in a reasonable amount of time. Although the field has not yet addressed concepts as complex as procedures, there are already some negative results with simpler concepts that bear on induction of procedures. Valiant (1984) presents strong evidence of the intractability of learning arbitrary boolean functions (a boolean function is an arbitrarily nested expression in propositional logic, containing just AND, OR and NOT). The class of procedures subsumes the class of boolean functions (because a sequence of actions is like an AND expression and a conditional branch is like an OR expression). So Valiant's evidence implies that procedure learning is also intractable.

However, Valiant concludes: "If the class of learnable concepts is as severely limited as suggested by our results, then it would follow that the only way of teaching more complicated concepts is to build them up from such simple ones. Thus a good teacher would have identify, name and sequence these intermediate concepts in the manner of a programmer." (Valiant, 1984, pg. 1135) Rivest and Sloan (1988) point out that having the teacher actually identify, name and sequence the subconcepts makes learning easy, but it places a great burden on the teacher. They present an algorithm that eases the load on the teacher but still insures successful learning. The algorithm can learn any concept representable as a boolean function, with the help of a teacher who breaks the concept into subconcepts and teach one subconcept per lesson, where a subconcept corresponds to a conjunction or disjunction in the boolean expression. This is based, of course, on a type of felicity condition that is quite similar to the one disjunct per lesson assumption.

In short, there are at least three general perspectives that offer explanations for the one-disjunct-per-lesson convention:

- The cognitive perspective: The student's mental processes for comprehending lesson material must do more work in order to correctly understand a lesson with more than one disjunct. If this work is not done or not done properly, misconceptions may develop.
- The social perspective: The classroom is a society that has persisted long enough to develop its own special conventions for facilitating social discourse, called felicity conditions. One disjunct per lesson is a felicity condition.

- The computational perspective: Learning from examples is governed by mathematical laws that determine the amount of computation necessary for success given the richness of the information accompanying the examples. A significant decrease in computation results from organizing a sequence of examples into lessons in such a way that only one disjunct is introduced per lesson.

These three perspective should not be considered as alternative hypotheses to be split by experimentation, but rather as mutually supportive.

The hypothesis that deserves testing is whether or not the one-disjunct-per-lesson convention actually does facilitate learning. This is an empirical question. The simplest test would be a two condition training experiment, where one condition's training material violates the one-disjunct-per-lesson convention and the other condition's training material obeys it. This paper reports the results of two such experiments.

All three perspectives sanction the one-disjunct-per-lesson convention, but they give no explicit predictions about what students will do if forced to learn from material that violates the convention. Thus, the experiments were exploratory. The basic idea was simply to videotape the students as they learned, and see if meaningful measures could be derived post hoc from transcripts of the training sessions. Protocol analysis has often been used successfully for exploratory studies of complex tasks (Ericsson & Simon, 1984).

Another design consideration was ecological validity. One of the strong points of the earlier study (VanLehn, 19??) is that it used real teachers teaching real students. In this experiment, an unusual curriculum was to be presented. Because it would not be ethical to run a mock classroom and risk "graduating" students with severe misconceptions, one-on-one tutoring was used. This would allow the tutor to correct, during the final session, any misconceptions that were acquired during the earlier sessions.

The need for verbal protocols suggested using older subjects than the second graders, who would be the natural choice if the subject material were subtraction, which was the chief subject matter in the earlier study. Thus, multiplication was chosen as the subject matter, because it is taught in third, fourth and fifth grades.

The one-disjunct-per-lesson hypothesis depends crucially on what counts as a disjunct, which in turn depends on how the knowledge is represented. To find a plausible representation for

multiplication, the simulation model developed for subtraction data, Sierra, was run on a lesson sequence for multiplication. Due to technical difficulties, it could not complete the last lesson (see VanLehn, 1987, for discussion). However, a fairly complete representation was obtained. Table 1 presents the an informal, simplified rendition of it.

Main procedure

If the multiplier (i.e., the bottom row) consists of a single digit,
 then call subprocedure Single-Digit-Multiply(N) where N is the digit,
 else if the multiplier consist of a non-zero digit followed by zeros,
 then call the subprocedure xN0,
 else call the subprocedure xNN.

Subprocedure Single-Digit-Multiply (N)

For each digit M in the multiplicand (i.e., the top row),
 multiply M by N,
 then add in the carry from the previous multiply if any,
 then write down the units digit of the result,
 then set the carry to the tens digit, if any.

Subprocedure xN0

For each zero in the multiplier,
 write down a zero in the answer.
 Call subprocedure Single-digit-multiply(N)
 where N is multiplier's nonzero digit.

Subprocedure xNN

For each digit N in the multiplier,
 if N is zero, then skip it,
 else
 write a zero for each digits in the multiplier to the right of N,
 then call Single-digit-multiply(N).
 Add up the partial products just generated.

Table 1: The multiplication procedure used in the experiments

Disjunctively related subprocedures can be located in the table by finding conditional statements (the "if ... then ... else" statements). There are two conditional statements. The first one selects among the subprocedures Single-digit-multiply, xN0 and xNN. Single-digit-multiply is used for problems such as 123×6 , xN0 is used for problems like 123×50 and 123×500 , and xNN is used for problems like 123×45 , 123×456 , 123×407 and 123×450 . The second conditional statement is located inside the xNN subprocedure. The first line of the conditional has a simple subprocedure that skips over zeros occurring in the multiplier. This would be used, for instance, in solving 123×406 and 123×450 . For future reference, this subprocedure is called the skip-zero trick.

2. Experiment 1

The experiment has two main conditions, called 1D/L and 2D/L. The training material in the 1D/L condition obeys the one-disjunct-per-lesson convention, while the training material of the 2D/L condition violates it. Logically, either the xN0 or the xNN subprocedure can be taught first. Thus, presentation order (xNN/xN0 vs. xN0/xNN) was crossed with the main manipulation, yielding four conditions. However, none of the dependent variables showed any effects for presentation order. For simplicity in subsequent discussion of the design and results, the presentation order manipulation is ignored.

2.1. Materials

The training material consisted of a four-lesson sequence. In both conditions, the first two lessons reviewed how to do single digit multiplication, and the second two taught the new subprocedures, xN0 and xNN. Each lesson consisted of the following four sections:

Review	The student reviews the preceding lesson's material by working a five-problem set.
Examples	The tutor works two examples out in detail, while the student watches and asks questions.
Samples	Two problems are worked with the tutor holding the pencil while the student tells the tutor what to write.
Practice	The student works about two pages of exercises at his or her own pace, asking questions as needed. The tutor watches carefully, and interrupts whenever a nontrivial mistake is made.

In order to screen subjects from the experiment who already knew the target material, two "trick" problems were included in the review section of lesson 3. One was a xN0 problem and the other was a xNN problem. Subject who correctly solved these problems were dropped from the experiment. (The other subjects all exhibited systematic errors that can be explained (informally, at least) by the impasse-repair process (VanLehn, 19??; Brown & VanLehn, 1980; VanLehn, 1983).)

In lesson 4, the second page of exercises for the practice section was actually (unbeknownst to the student) a page of transfer problems whose multipliers had a mixture of zeros and non-zero digits in them (e.g., xN0N, xNN0N, xN00N, etc.). The purpose of these exercises was to see if students could invent the skip-zero trick on their own. The mathematical principles behind it had been presented as part of their instruction in the xN0 and xNN methods. If they really understood the two methods, they might be able to invent the skip-zero trick. Thus, the transfer section of lesson 4 was designed to test depth of understanding.

2.2. Subjects and methods

Subject acquisition and retention were problematic. The experiment needed subjects who were at a particular point in their schooling and who would be willing to come to the university for four one-hour sessions. Although subjects were paid for their participation (one silver dollar per lesson) and even chauffeured to the laboratory, it was still difficult to find volunteers. The experiment ended with only 8 subjects in the 1D/L condition and 7 in the 2D/L condition. Four subjects, one in each condition, came from academic families whose children were in a university nursery school. The rest of the subjects were recruited with newspaper advertisements.

Subjects were run individually. Sessions lasted between 45 minutes and an hour. When possible, sessions were scheduled on four consecutive days. However, in some cases the last session did not occur until two weeks after the first. Most subjects were run during the summer in order to avoid intrusion of their normal mathematics classes into the experimental teaching. The subjects who were run during the school year were asked what they were learning during school; according to the subjects, multiplication was not taught in school during the course of the experiment.

2.3. Results

The experiment was designed with no specific predictions about how learning would differ among the conditions, so all the sessions were videotaped and lessons 3 and 4 were transcribed.

Qualitatively, there was little apparent difference between conditions. All the students found the learning task non-trivial, but some students quickly assimilated the instruction and mastered the skill, while others never really understood the algorithm despite valiant efforts on their part and the tutor's. Qualitatively, it did not look like the differences in performance were caused by the conditions, but rather were caused by individual differences among the subjects.

In order to quantify the degree of confusion engendered by the training, the subject's errors were counted using their worksheets and the protocol transcripts. Facts errors, such as $3 \times 5 = 12$ or $7 + 9 = 13$, and errors in carrying were not counted since these skills were taught prior to the experiment.

Although these error counts could be used as the dependent measure, two aspects of the

experiment suggested using a more complicated measure. Although all the errors were corrected, either by the tutor or by the students themselves. Often, the tutor would simply point out the existence of the error with a word or gesture, and the student would correct it. Sometimes the tutor would give long explanations. Thus, these errors should be considered instances of communication between the tutor and the student, and not just signs of miscomprehension. However, some students tended to ask the tutor for help if they are unsure rather than make a mistake and have it corrected. Such questions were also be counted along with the errors, since they too are instances of communication caused by a lack of understanding.

Another consideration is that students worked at different rates, mostly because their familiarity with the multiplication facts varied widely. Thus, it would not make sense to compare error/question counts across subjects, since the faster subjects have more opportunities to make errors. In order to factor out the effects of varying speeds, we counted opportunities for errors as well as errors. For these purposes, an error categorization was developed. Table 2 lists the six categories used. For each category, the first line describes what the student should do, and the second line describes the modal error for that category. Thus, if the following solution is generated by a student,

$$\begin{array}{r}
 123 \\
 \times 201 \\
 \hline
 123 \\
 0000 \\
 + 24600 \\
 \hline
 24723
 \end{array}$$

then the count of opportunities for category X would be increased by one, because the student had the opportunity to exercise the skip over the zero in the multiplier. However, the student did not use the skip-zero trick, so the error count for category X is also increased by one.

The term "spacer" in table 2 refers to the zeros that are placed on the right end of a partial product's row. Although not all multiplication algorithms use spacers, the one taught in the experiment did.

The number of errors and questions of each type were divided by the number of opportunities of that type. This calculation yields a rate, which is similar to an error rate, except that it includes questions as well as errors. It will be called a *confusion rate*. Confusion rate is the dependent measure in the results reported below.

-
- | | |
|---|---|
| X | Skipping a multiplier digit if it is zero
Multiplying the multiplicand by zero, generating a row of zeros |
| S | Multiplying the multiplicand by a non-zero multiplier digit
Skipping a non-zero multiplier digit |
| Z | Remembering to write the spacer zeros of a row
Forgetting to write the spacer zeros |
| N | Writing the right number of spacer zeros
Writing some spacer zeros, but the wrong number of them |
| R | Moving to the next row before writing the next partial product
Concatenating a partial product to the left end of the previous one |
| P | Remembering to add up the partial products
Failing to add up the partial products |

Table 2: Protocol coding categories.

If any difference between the 1D/L and the 2D/L conditions is found, it could be attributed to differences in the *amount* of teaching given to the subjects, and not the organization of it into lessons. To check this, the words uttered by the tutor during the initial instruction on the algorithm were counted (the Examples sections of lessons 3 and 4 for 1D/L; the Example section of lesson 3 for 2D/L). The means (664 words for 1D/L and 520 words for 2D/L) were not significantly different.

Table 3 shows the main results, the confusion rates per category and per section. The first two columns of figures shows the confusion rates for the introductory sections of the lessons. For the 1D/L condition, the introductory sections were the Example and Sample sections of lessons 3 and 4. For the 2D/L condition, the introductory sections were Example and Sample sections of lesson 3. The second two columns show the Practice sections of lessons 3 and 4. The last two columns shows the Transfer section of lesson 4.

Mean confusion rates for single categories are shown in the top part of the table. The lower part shows combinations of related categories. These combinations are intended to indicate how the results would come out if larger categories had been used. For instance, the combination of Z and N categories measures the rate of all confusions that involve spacers. The combination of S and X categories measures includes all confusions that mix up the xNN method with the xN0 method, on a narrow interpretation of "mix up." The combination of S, X, R and P categories is a

Type	Intro. 1D/L	Intro. 2D/L	Prac. 1D/L	Prac. 2D/L	Trans. 1D/L	Trans. 2D/L
S. Skip non-zero digit	.276	.221	.085	.114	.426*	.063*
X. Omit skipping zero digit	.039	.142	.000	.119	.673	.300
Z. Omit spacers	.264	.148	.024?	.169?	.202	.262
N. Wrong number of spacers	.073	.033	.016?	.113?	.232	.336
R. Concatenate rows	.067	.040	.068	.146	.150	.032
P. Omit adding up	.144	.071	.052	.074	.146*	.000*
Types S and X	.186	.194	.047	.116	.529**	.155**
Types Z and N	.147?	.079?	.019?	.135?	.223	.312
Types S, X, R, and P	.159	.152	.051	.118?	.381**	.103**
All types	.154	.193	.037?	.113?	.309	.197

Table 3: Mean confusion rates during the introductory, practice and transfer sections.

** = $p < .01$, * = $p < .05$, ? = $p < .10$.

broader interpretation of the notion of mixing up the xNN method with the xN0 method. Combination confusion rates are calculated in the same way as a regular confusion rate, by dividing the number of confusions of that (larger) type by the number of opportunities for that type of confusion to occur. The difference between the means of two conditions may be significant for a combination category even when the difference for the means of its constituent categories is never significant. This can occur when the category combines subjects who are high in one type of error with subjects who are high in another type; this reduces the variance, leading to significance.

In both the introduction and practice sections of the training, T-tests indicate that none of the mean confusion rates were significantly different. The overall confusion rate, combining all categories in both sections, was .090 for the 1D/L condition and .112 for the 2D/L. This difference was not significant.

However, there were two marginally significant differences. The first involves confusions about spacers (category Z). In the introductory sections, the 1D/L students were more confused than the 2D/L students, while in the practice sections, the opposite trend is found. Apparently, both conditions suffered confusions about spacers, but the 1D/L condition exhibited their confusions earlier. Indeed, when the confusion counts from the introductory and practice sections are combined, the difference between the means disappears.

There is a tendency in the practice sections for the 2D/L subjects to exhibit more confusion than the 1D/L subjects about when to use the xN0 method. They tended to use the xNN method

even when the multiplier had the xN0 form (category X). This difference is evident in the introductory sections as well. When the counts for the introductory and practice sections are combined, the difference is marginally significant (.016 for 1D/L versus .129 for 2D/L; $p < .114$).

In the transfer section (the last two columns of the table), there was a strong effect due to the experimental manipulation. Although the two conditions did not differ on the number of spacer errors (categories Z and N), they differed significantly on the other types (categories S, X, R and P) which presumably are due to mixing up the xNN and xN0 method. However, the trend was in the opposite direction from that predicted by the one-disjunct-per-lesson hypothesis. The 1D/L condition had a confusion rate that was three times larger than the 2D/L condition.

2.4. Discussion

The one-disjunct-per-lesson hypothesis predicts that the 1D/L students would be less confused by their training than the 2D/L students. The data did not confirm this prediction, although the confusion rates for the X-type errors were in the right direction for the introductory and practice sections.

A major effect, which was unexpected, is that the 2D/L students are better at solving problems that combine the xNN and xN0 methods. These problems, which first occur in the transfer section, have multipliers of the form xN0N, xN0NN, xN00N, and so on. The preponderance of errors of types S, X, R and P indicates that 1D/L students are mixing up the xNN and xN0 methods. The S errors indicate that they are using the xN0 method (or at least the skip-digit part of it) when the xNN method is appropriate. The X errors indicate that they are using the xNN method on zero multiplier digits, which causes them to generate a row of zeros. If they had used a shifted-over version of the xN0 method, they would have avoided this. In short, it seems that the 1D/L students have not learned to discriminate the conditions under which the two methods are appropriate, and this caused them to make three times as many errors during the transfer section as the 2D/L students.

One possible explanation of the main effect is that the 2D/L students were trained with mixed drill, where different methods are used on different problems in the same lesson, whereas the 1D/L students were trained with homogeneous drill, where the same method is used on every problem in a lesson. Thus, the 2D/L students had to learn *when* to apply the two methods, whereas the 1D/L

students could simply use the same method as they used on the previous problem, and never bother to induce the applicability conditions of the method. Thus, one would expect the mixed drill to facilitate performance on the transfer sections, where applicability of methods is important, and this is indeed the observed result.

3. Experiment 2

A possible explanation for the lack of an effect during training for the 2D/L condition is that students understood a "lesson" to be a single example, rather than a whole session/lesson. It is logically possible to define a lesson to be as small as a single example or even a part of an example. In the subtraction studies (VanLehn, 19??) and this study, a lesson was defined to be the material taught in one class session. However, this might not be the definition of lesson that students use.¹ That is, if students had a felicity condition of one-disjunct-per-example, then neither the 1D/L nor the 2D/L curricula of experiment 1 would violate their felicity condition. This would explain why the confusion rates are not easily distinguished across conditions.

To explore this possibility, training material was constructed that attempts to teach two disjuncts in the same example and thus violate the felicity condition of students using a one-disjunct-per-example convention. This material cannot use the xNN and xN0 methods as its disjuncts, because they require different types of problems. So two new disjuncts had to be used. Unfortunately, this makes the performance on this training material incomparable with the performance of the 1D/L and 2D/L training material, because it teaches a different subject matter. Although this training material was run at the same time as the other two, fewer subjects were run and the analysis is different, so it is described as a separate experiment.

3.1. Subjects and methods

Four subjects were run, using the same subject selection procedures as in experiment 1. The subjects were run in the same manner, in four sessions of tutorial instruction, with each session divided into four sections. As in the first experiment, the first two lessons reviewed single digit multiplication. The new material was introduced in lesson 3. The difference in procedure between this experiment and experiment 1 is the subject matter of the lessons 3 and 4.

3.2. Materials

The material in lessons 3 and 4 was designed to teach the skip-zero trick (see table 1). Except for the transfer section at the very end of the fourth session, students saw only problems whose multipliers had the form $xN0N$, $xN00N$, $xNN0N$ and $xN0NN$. On the initial example, which had the form $xN0N$, the tutor would verbally describe the xNN method as part of the explanation of multiplication of the multiplicand by the units digit. The rationale for the skip-zero step is explained during the multiplication by the tens digit, which is a zero. Thus, the students hear about two distinct methods for producing partial products during the initial example. They also see instances of both disjuncts. However, the instance of the skip-zero disjunct occurs on a different subproblem (i.e., the tens-digit multiply) than the instances of the regular digit multiplication. Table 4 illustrates this with the relevant section of one of the protocols. Notice how the tutor combines the explanations for the xNN method and the skip-zero trick.

Tutor:	This is how you do these. [Displays worked example: 203×102 .] What you were telling me first was pretty much right. First you look at the number in the ones column. [Points to the 2] You pretend the other two aren't even there.
Subject:	OK.
Tutor:	And then you multiply out. You say, 2 times 3 is 6, 2 times 0 is 0, 2 times 2 is 4. OK?
Subject:	Uh huh.
Tutor:	That's the first step. Ok. Now, the way it works is this number gets its own row, which you just made, and now when you move on to the next number, it gets its own row, too. You make another row. OK? But in this case, since the tens number equals 0 [Points to the zero in the multiplier], you get to skip the row. Because, see, if you multiplied out the tens number, you'd say 0 times 3 is 0, 0 times 0 is 0, 0 times 2 is 0; and you'd have a whole bunch of zeros.
Subject:	Uh huh.
Tutor:	Right. So you can skip that row. OK. So you can move on the the hundreds number [Points to the 1 in the hundreds place of the multiplier]. Ok, so now you're just looking at the hundreds number. And the only trick about this--and this is something that you just have to remember--when you do the ones number, you just do it like this. Ok, when you move on out to the tens, what you'd do if we were going to work it out is you'd say, I have 0 ones. First, you'd put a 0 down.
Subject:	Yeah.
Tutor:	Ok. Then you'd multiply out: 0 times 3 is 0, 0 times 0 is 0, 0 times 2 is 0. Now, when you move on to the hundreds, you have 0 ones and 0 tens. So you put down two zeros first. OK? And then you multiply out. You say 1 times 3 is 3, 1 times 0 is 0, 1 times 2 is 2. OK?
Subject:	Uh huh.
Tutor:	No.
Subject:	I don't understand.
etc.	

Table 4: A fragment of protocol showing the initial training on the $xN0N$ method

As in experiment 1, a transfer section was included at the end of lesson 4. The transfer section used problems whose multiplier were of the form xNN and $xN0$.

3.3. Results

The protocols were transcribed and coded as before. Table 5 shows the resulting confusion rates for all three sections.

Type	Introductory	Practice	Transfer
S. Skip non-zero digit	.270	.093	.088
X. Omit skipping zero digit	.219	.099	.050
Z. Omit spacers	.333	.157	.192
N. Wrong number of spacers	.059	.061	.171
R. Concatenate rows	.119	.089	.083
P. Omit adding up	.161	.031	.050
Types S and X	.245	.096	.069
Types Z and N	.196	.109	.182
Types S, X, R, and P	.192	.078	.068
All types	.271	.103	.124

Table 5: Mean confusion rates during the 2D/E training

3.4. Discussion

Because the 2D/E training taught different material than the 1D/L and 2D/L training, it would not be meaningful to compare their confusion rates statistically. However, it is clear that the overall confusion rates in the three sections of the 2D/E training (Introductory: .271; Practice: .103; Transfer: .124) were of the same order of magnitude as the overall confusion rates for the other two conditions (Introductory: .154 and .193; Practice: .037 and .113; Transfer: .309 and .197; for 1D/L and 2D/L, respectively). The expected aberrant behavior due to violation of a felicity condition did not seem to occur. This confirms the subjective impression one has on viewing the video tapes that the students in this experiment acted just about the same as the students in experiment one.

4. General Discussion

The results of experiments one and two show that violating the one-disjunct-per-lesson (or one-disjunct-per-example) felicity condition does not lead to unusual or extraordinary learning behavior. The confusion rate data indicate that the differences between conditions were at best marginally significant, except in the transfer section of experiment one.

In the transfer section of experiment one, the 1D/L students made three times as many errors as students in the 2D/L condition. This finding is consistent with the hypothesis that mixed drill, such as that found in the 2D/L condition, requires learning how to discriminate problems requiring one method from problems requiring another, but that homogeneous drill, such as that found in the 1D/L condition, does not cause students to learn such discrimination information. Thus, the 2D/L students do better on the transfer section, because that section's exercises require choosing which method to use on each problem.

It may seem that the predicted effects of the felicity conditions could exist but be hidden by the mixed drill effect. However, this cannot be the case. The one-disjunct-per-lesson hypothesis predicts that the 2D/L students should be more confused by the introductory and practice sections than the 1D/L students. For the mixed drill effect to hide the predicted effects, it would have to reduce confusion during those two sections. However, the mixed drill of the 2D/L condition should be, if anything, more confusing during the introductory and practice sections than the homogeneous drill of the 1D/L condition, because the mixed drill students need to learn more than the homogeneous drill students. The 2D/L students need to learn both *how* to do the xNN and xN0 methods and *when* to do them, whereas the 1D/L students only need to learn *how* to do the methods. Thus, the mixed drill effect should add to the confusion of the 2D/L students, rather than subtracting from it. Since the 2D/L students were not significantly more confused than the 1D/L students during the introductory and practice sections, neither the felicity conditions nor the mixed drill training seemed to have a profound effect in those sections.

A possible explanation for the lack of a felicity condition effect is that not enough subjects were run for the trend in the data to achieve statistical significance. The overall confusion rate in experiment one (excluding the transfer section of lesson four, due to the mixed drill effect) is .090 for the 1D/L condition versus .112 for 2D/L -- a small trend that is rendered nonsignificant by the high variance among the subjects. However, my subjective impression on viewing the video tapes is that if there is a felicity condition effect in this experiment, it is quite small in comparison to the vast individual differences among subjects. This is just what the statistics say, too.

The lack of a strong felicity condition effect is probably due to the use of one-on-one tutoring instead of classroom instruction. It could be that the students in the 2D/L and 2D/E conditions could actually have been quite confused by the instruction, but the tutor was able to remedy their

confusion so quickly that it does not show up in the confusion rates. Indeed, the tutor gave long verbal explanations of the two multiplication methods as well as answering any questions posed by the student. Moreover, whenever the student made a serious mistake, the tutor would interrupt, correct the mistake, and explain the correction. This immediate, rich feedback means that most confusions introduced by the lesson material were quickly remediated.

The only kind of misconception that could escape this teaching method would be one where an incorrect piece of knowledge happened to yield error-free performance on training material. This may be what happened with the students in the 1D/L condition. They may have adopted the heuristic of always choosing the method that they used on the previous problem. During training, this buggy heuristic yielded correct solutions, so the tutor had no cause to interrupt and remediate. Consequently, the mistaken heuristic persisted into the transfer section, where it caused the 1D/L students to commit many errors.

On this view, the beneficial effects of one-on-one tutoring were so powerful (cf. Bloom, 1984) that they wiped out any confusions that the violation of felicity conditions might have caused, but allowed the confusing effect of mixed drill to come through unscathed.

If this explanation of the lack of a felicity condition effect is correct, then one-disjunct-per-lesson may still have a large effect in non-tutorial situations, such as classroom teaching. If the lesson material introduces a confusion, students may harbor it for minutes, days, or years before it is detected and remediated. If so, then this would explain why curriculum designers tend to obey one-disjunct-per-lesson, because in the classroom context, it really does make a difference how many disjuncts are packed into a lesson.

A second explanation for the lack of an effect is that the tutorial mode could be so effective at communicating ideas that the added clarity imparted by a well-structured training sequence is not needed. As discussed in the introduction, the main purpose of the one-disjunct-per-lesson convention might be to use lesson boundaries to tell the student when to add a new disjunct. However, another way to communicate that a new disjunct should be added is simply to announce, "Now I'm going to show you something new." The tutor in these experiments always prefaced the introduction of disjuncts with such a statement, even in the 2D/E training. Sometimes the statement was subtle. For instance, in table 4, the tutor said only, "But in this case," before

introducing the discussion of the skip-zero disjunct. Note that this announcement occurs in the middle of solving 203×102 . In a classroom situation, such a subtle preface might be easily overlooked. This suggests that the observed tendency of curriculum designers to place disjunct introductions at the beginning of lessons might be a way of providing a forum for the teacher's announcement that "Today, we're going to learn something new." Such a forum may be superfluous in the tutorial setting but important in the classroom. This provides a second explanation as to why the felicity condition manipulation had no effect in the experiments reported here.

To put it differently, these experiments are consistent with a general felicity condition that could be paraphrased as a command to students: "Don't disjoin unless I tell you too. Generalize instead." For classrooms, the "telling" happens to occur at the beginning of lessons, but this is a rhetorical convenience only.

References

- Anderson, J. R., Farrell, R., & Saurers, R. (1984). Learning to program in LISP. *Cognitive Science*, 8, 87-129.
- Austin, J. L. (1962). *How to do things with words*. New York, NY: Oxford University Press.
- Badre, N. A. (1972). *Computer learning from English text*. Berkeley, CA: University of California at Berkeley, Electronic Research Laboratory.
- Bloom, B.S. (1984). The two sigma effect: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 3-16.
- Brown, J. S. & VanLehn, K. (1980). Repair Theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.
- Chi, M.T.H., Bassok, M., Lewis, M., Reimann, P. & Glaser, R. (in press, 19??). Learning problem solving skills from studying examples. *Cognitive Science*, .
- Ericsson, K.A. & Simon, H.A. (1984). *Protocol Analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- LeFevre, J. & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction*, 3(1), 1-30.
- Pirolli, P.L. & Anderson, J.R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, 39(2), 240-272.
- Rivest, R. & Sloan, R. (1988). Learning complicated concepts reliably and usefully. In MIT (Ed.), *COLT 88: Proceedings of the first workshop on Computational Learning Theory*. ??? : ??? To appear in Proceedings of AAAI-88.
- Valiant, L.G. (1984). A theory of the learnable. *Communications of the ACM*, 27(11), 1134-1142.
- VanLehn, K. (1983). Human skill acquisition: Theory, model and psychological validation. In *Proceedings of AAAI-83*. Los Altos, CA: Morgan Kaufmann,
- VanLehn, K. (1986). Arithmetic procedures are induced from examples. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics*. Hillsdale, NJ: Erlbaum.
- VanLehn, K. (1987). Learning one subprocedure per lesson. *Artificial Intelligence*, 31(1), 1-40.
- VanLehn, K. (in press, 19??). *Mind Bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

Notes

¹Indeed, older students would have to use smaller lessons in order to make sense of, say high school algebra because algebra texts sometimes introduce several subprocedures (e.g. algebraic transformations) in material designed to be covered in one class period. Fortunately for the one-disjunct-per-lesson hypothesis, such material is often organized as a sequence of blocks of materials, with clearly evident boundaries between them.