Assessing Schematic Knowledge of Introductory Probability Theory

Damian P. Birney[1]

University of Sydney

Gerard J. Fogarty

and

Ashley Plank

University of Southern Queensland

Key Words: assessing schematic knowledge, text-editing, statistical probability theory

[1] Author's address:

Damian P. Birney, School of Psychology, University of Sydney, NSW, 2006, Email: damianb@psych.usyd.edu.au; Tel: +61-2-9351-4236; Fax: +61-2-9036-5223

Running Head: Schematic Knowledge of Probability Theory

Abstract

The ability to identify schematic knowledge is an important goal for both assessment and instruction. In the current paper, schematic knowledge of statistical probability theory is explored from the declarative-procedural framework using multiple methods of assessment. A sample of 90 undergraduate introductory statistics students was required to *classify* 10 pairs of probability problems as similar or different; to identify whether 15 problems contained sufficient, irrelevant, or missing information (*text-edit*); and to *solve* 10 additional problems. The complexity of the schema on which the problems were based was also manipulated. Detailed analyses compared text-editing and solution accuracy as a function of text-editing category and schema complexity. Results showed that text-editing tends to be easier than solution and differentially sensitive to schema complexity. While text-editing and classification were correlated with solution, only text-editing problems with missing information uniquely predicted success. In light of previous research these results suggest that text-editing is suitable for supplementing the assessment of schematic knowledge in development.

Assessing Schematic Knowledge of Introductory Probability Theory

Knowledge is many things to many people. de Jong and Ferguson-Hessler (1996) have commented that knowledge has been conceptualised as domain-general and domain-specific, concrete and abstract, implicit and explicit, formal and informal, elaborated and compiled, declarative and proceduralized, conceptual and procedural, unstructured and structured, tacit or inert, strategic, situated, schematic, as knowledge-acquisition knowledge, as metaknowledge – and the list goes on. It would seem that *knowledge*, like intelligence, is a rather nebulous construct. However, within this quagmire a small number of approaches to conceptualising knowledge have received almost consensual support. The declarative-procedural framework is one such approach for which evidence has amassed across numerous domains (Anderson & Schunn, 2000; Benaroch, 2001; Cohen, Poldrack, & Eichenbaum, 1997; Kirasic, Allen, Dobson, & Binder, 1996; Low & Over, 1992; Rittle-Johnson, Siegler, & Alibali, 2001). How we conceptualize knowledge not only determines the nature of our instructional strategies but also how we assess it. In the current study we adopt the general declarative-procedural framework to explore the assessment of *schematic* knowledge of probability theory. This element of statistics is particularly difficult for undergraduate students to master (Konold, 1995) and there are ongoing calls for more innovative ways to assess statistical knowledge (Garfield, 1994). Schematic knowledge is highly context specific (Quilici & Mayer, 2002). It contains declarative knowledge (also referred to as conceptual knowledge, de Jong & Ferguson-Hessler, 1996; Rittle-Johnson et al., 2001) and procedural knowledge, in addition to *situational* knowledge that provide cues to when knowledge should be used (Cheng & Holyoak, 1985; de Jong & Ferguson-Hessler, 1996). A traditional method used to assess schematic knowledge is to have students classify pairs of problems in terms of structural relatedness. We compare this to a relatively new method that has received little attention in the literature called text-editing. Before we describe the text-editing task, a brief overview of the declarative-procedural-schematic framework is presented.

Declarative knowledge.

Declarative knowledge is commonly considered as knowing-what, and *procedural knowledge* as knowing-how. The declarative-procedural framework has been at the heart of numerous information-processing theories of knowledge-acquisition (e.g., Anderson, 1982; Anderson & Milson, 1989; Anderson & Neves, 1981; Anderson & Schunn, 2000). During the early stages of skill acquisition, declarative information can be used with general problem-solving procedures in an interpretative way (Anderson, 1990). For example, a student learning to solve *area-of-rectangle* problems in a mathematics course may initially encode declarative information relating the length of adjacent sides of a rectangle to the surface area. The student is likely to have had some experience solving other mathematics problems and will also have access to general problem-solving strategies used in everyday reasoning. These additional experiences can be brought to bear on the newly acquired declarative information about rectangles to facilitate success without the student necessarily possessing specific solution *strategies* or *procedures* (Anderson, 1982; Anderson & Neves, 1981; Rumelhart & Norman, 1981).

Two examples of general problem-solving approaches that have received considerable attention in the literature are *working forward* and *working backward* strategies[1]. An example of a working forward reasoning strategy would be making a supposition that a particular formula for solution is appropriate from the start, and then following the ensuing inferences through to their logical end (Byrne & Handley, 1997; Rips, 1989). An example of a working backward strategy is the *means-ends approach*, which is essentially a goal directed, search-oriented process frequently used by novices (e.g., Gick, 1986). The idea behind this procedure is that differences between the goal-state specified by the problem and successive problem states obtained during the course of solution are analysed in order to choose the next best move or sub-goal (Sweller & Levine, 1982). While means-ends analysis is an efficient and flexible problem-solving strategy in novel situations where specific procedures or schemas are lacking, it is resource-demanding because a comprehensive search of memory is necessary at each sub-goal (Gick, 1986; Sweller, 1988, 1989). In fact, general problem-solving procedures can make such high demands on working memory resources that they can actually interfere with learning (Van Gog, Paas, & Van Merriënboer, 2004). For instance, Sweller and Levine (1982) showed that maze learning was inhibited by means-ends approaches. Similar inhibitory effects on learning have been shown in algebraic word problems (e.g., Cooper & Sweller, 1987; Ward & Sweller, 1990; Zhu & Simon, 1987). The inference from these studies is that general problem-solving consumes working memory resources that would be better spent on learning the structure of the task at hand (Sweller, 1988, 1989, 1993; Sweller, Chandler, Tierney, & Cooper, 1990; Sweller, Mawer, & Howe, 1982; Woltz, 1988).

Procedural knowledge and learning.

A common theme in many cognitive learning theories is that people construct new knowledge by “assimilating” their current experience with what they already know or believe to be true about the world (Chen, 1999; Cook, 2001; Gick & Holyoak, 1983; Piaget, 1950). As experience in a domain increases, related declarative facts and specific solution strategies are linked – they become proceduralized. Instruction is often premised on this understanding. That is, while proceduralization means that the use of knowledge becomes less dependent on domain-general reasoning, proceduralization has also been shown to facilitate further learning (e.g., Rittle-Johnson et al., 2001). Two aspects related to this are problem structure and context. We consider each briefly.

Structure: One of the main objectives of instruction is to help students to focus on the structure of a domain and to draw their attention away from the more transient surface details of a particular problem. For instance, altering the format of instructions of learning tasks to be less goal oriented has been shown to facilitate learning because it encourages learners to be more sensitive to the structure of the task (Chandler & Sweller, 1991; Sweller, 1993; Sweller et al., 1990; Sweller & Levine, 1982; Van Gog et al., 2004) and less focused on surface details (Quilici & Mayer, 1996, 2002). Imagery has also been used to improve understanding of mathematical structure. An important feature of visual-spatial presentation on knowledge acquisition has to do with how elaborated learning can facilitate the recall of important structural features of problems. However not all types of visual-spatial representations have been shown to be equally effective in improving mathematical achievement (Novick, Hurley, & Francis, 1999). Hegarty and Kozhevnikov (1999) suggested that the use of schematic imagery (i.e., pure relationships depicted in a visual-spatial scheme devoid of concrete images) was positively correlated with mathematical performance, but that there was a negative correlation with use of pictorial imagery (i.e., concrete images in the mind). Pictorial imagery was associated with poorer performance because it takes the problem solver's attention away from the problem structure.

Context: Presenting problems in a familiar context has been shown to enhance knowledge acquisition and performance while at the same time reducing working memory demand (Quilici & Mayer, 1996, 2002). For example, Carraher, Carraher, and Schliemann (1985) studied Brazilian children who, for economic reasons, often worked as street vendors (see also, Nuñes, 1994). Mathematical problems were either embedded in real-life situations, where appropriate *procedures* and strategies for using declarative knowledge were in place (i.e., vending), or in an academic context, where, for this population, procedures were less developed. The children's performance was significantly better in the former problems.

Schematic knowledge.

A more extended conceptualisation of the organisation of knowledge can be found in the concept of a schema. Schemas have been defined in many ways. Within mathematics, Cooper and Sweller (1987, p. 348) defined a schema as "...a construct that allows problem solvers to group problems into categories in which the problems in each category require similar solutions". Similarly, Sweller (1989, p. 458) defined a schema as a "...cognitive construct that permits problem solvers to recognise problems as belonging to a particular category requiring particular moves for solution". Furthermore, schemas are useful because they serve to categorise knowledge according to the way in which it will be used (Sweller, Van Merriënboer, & Paas, 1998). Although originating from slightly different lines of research (Chi, Fletovich, & Glaser, 1981; Sweller et al., 1998), schema theory fits closely with Anderson's (1982; 1990) declarative-procedural framework. A common finding in the expert-novice literature has been that experts' organization of knowledge is qualitatively different from that of novices and that this difference is a function of experts' greater access to elaborated schemas (Chi et al., 1981; de Jong & Ferguson-Hessler, 1986) that develop with experience in the problem domain (Sweller et al., 1998).

Contextualized schematic knowledge reflects a deep structural understanding of the domain, but this high contextualization can also come at a cost. It may place constraints on the availability of certain information and in some cases can produce near-transfer failure – the unexpected failure to transfer knowledge from one context to a similar one in the *same* domain (Woltz, Gardner, & Gyll, 2000). The fact that people can solve mathematics problems presented in a familiar practical context but fail dismally when only the context is changed (Carraher et al., 1985), is a clear example of one of the disadvantages of contextualized knowledge. However, because schemas are highly contextualized, under most circumstances they can be accessed directly without the separate facts and procedures having to be retrieved, interpreted, and assessed for appropriateness piece by piece. The end result is generally a faster more efficient processing system (Anderson, 1982, 1990; Anderson & Neves, 1981).

There is evidence to suggest that the development of knowledge and expertise in a domain does not necessarily move from declarative, to procedural, to schematic in a smooth monotonic fashion (e.g., Rittle-Johnson et al., 2001) and that declarative and procedural knowledge exists side-by-side with schematic knowledge (de Jong & Ferguson-Hessler, 1986). For instance, Savelsbergh, de Jong, and Ferguson-Hessler (2002) suggest that even though concepts in many domains (especially mathematically based domains) can be ordered hierarchically from concrete to abstract, it is often unclear what is abstract and what is concrete, and how such a distinction might change from one individual to another. Nevertheless, there is evidence to suggest that depth of knowledge does improve as knowledge becomes proceduralized, and that expertise increases with the acquisition and development of schemas (Chi et al., 1981; Chi, Glaser, & Rees, 1982). Mayer (1987) argued that developing an understanding of problem structure should be the primary instructional goal in teaching mathematics. Schematized knowledge reduces cognitive load, freeing resources for the acquisition of problem structure, which in turn has been repeatedly shown to facilitate near and far transfer of skills (e.g., Sweller et al., 1998; Van Gog et al., 2004). The ability to identify schematic knowledge is therefore a worthwhile goal for both assessment and instruction (Ngu, Low, & Sweller, 2002; Quilici & Mayer, 2002).

Assessing Schematic Knowledge

Various techniques have been used to assess schematic knowledge. Students have been asked to recall details of problems they have either seen or solved previously (Low & Over, 1989, 1990). Recall of only surface details of problems has been shown to indicate poor schematic knowledge, whereas recall for structural details indicates that schemas have been used (Low & Over, 1990; Ngu et al., 2002). Another method of assessing schematic knowledge is to have students categorise problems in terms of their perceived similarities. As with recall, it is assumed that persons with poor schematic knowledge of the domain will sort by surface similarities, whereas persons with better knowledge will consider deeper structural components (Littlefield & Rieser, 1993; Low & Over, 1989, 1990, 1992; Schoenfeld & Herrmann, 1982). The criterion often used to validate these techniques is the student's ability to actually solve problems.

The approach to assessing schematic knowledge that we used in this study is called *text-editing* (Low & Over, 1989, 1990, 1992, 1993; Low, Over, Doolan, & Michell, 1994; Ngu et al., 2002). The idea is that the ability to identify what information from the text of a problem should be used, in what sequence, and through what operations, reflects an understanding of the problem structure and is therefore evidence of appropriate schematic knowledge. Text-editing requires students to isolate the necessary and sufficient components needed to solve word problems. The problems are designed so that they fit into one of three mutually exclusive categories: (1) there is sufficient information given to solve the problem; (2) there is sufficient information plus irrelevant information; and (3) information required for problem solution is missing. In the last two cases, students are also required to identify the irrelevant or missing information. To illustrate the nature of text-editing, consider the following problem which was used by Low et al. (1994) to train students in text-editing *area-*

of-rectangle problems.

> *The length of a rectangular window pane is twice its width. The area of the pane is 98 cm². What are the dimensions of the pane?*

If we are to consider the schema required to solve this problem, knowledge that the area is equal to the length multiplied by the width would be one necessary component (Low & Over, 1989). If the width = a and the length = $2a$, as given in the problem, then retrieval of an appropriate *area-of-rectangle* schema that integrates potential solution procedures and necessary declarative facts would facilitate solution. It is not a requirement of text-editing to actually solve the problem, however in this case there is sufficient information provided to do so. If the student does not possess an appropriate schema to solve the problem, the probability of making an error in the text-editing task is increased.

Low and Over (1989) were able to reinforce their claims for the value of text-editing as a direct measure of schematic knowledge by demonstrating that performance on text-editing was strongly correlated with problem solution. They demonstrated this in two ways: Firstly by showing that text-editing predicted as much as 90 percent of the variance in subsequent problem solution; and secondly by showing that failure in the text-editing section of their experimental procedure was almost invariably followed by failure in the problem solution section. While text-editing seems to be a necessary condition for problem solution, it is not sufficient. Various studies have shown that a significant proportion of students were able to correctly text-edit problems but then failed solution (Low & Over, 1989, 1990, 1992) – we explored this finding further in the current study. Low and Over (1990) went on to show that text-editing is correlated with recall memory for algebraic word problems and discrimination of whether problems are similar or different. In this second study, text-editing scores were also correlated with tests of general mathematical ability, even after allowance was made for contributions from verbal ability. The authors concluded that because the task requires an understanding of problem structure, text-editing can be taken as a measure of schematic knowledge. Previous research by Low and colleagues indicates that text-editing is easiest for problems where there is sufficient information and most difficult where the problems contain irrelevant information. The extent that this pattern of results generalizes to probability problems is an empirical question.

The successful application of text-editing techniques in mathematics has paved the way for its application in other fields. Ngu et al. (2002) examined its usefulness as a training aid in the field of chemistry and concluded that its effectiveness depended on the learning materials. Our particular interest in this study was in the field of statistics. There are some obvious parallels between the fields of mathematics and statistics. The format in which statistics problems are encountered by students is typically similar to the format of algebraic word problems used in text-editing research (Low & Over, 1989, 1990; Mayer, 1982). Furthermore, statistical probability problems are based on a number of related but distinct classes that require different formulas and procedures for solution. As detailed in the Method section, the text-editing problems used in the current study are drawn from five problem classes (see Table 1). These classes can be differentiated on the sophistication of the schema involved. For instance, solving probability problems using the addition rule when events are not mutually exclusive (Class A-NM in Table 1) is a more complex instantiation of the basic addition rule with mutually exclusive events (Class A-M). Similarly, Class M-NI involved a more complex instantiation of the multiplication rule than Class M-I. These classifications also tend to coincide with typical teaching approaches in these areas – simpler, more specific problem classes are often taught first followed by instruction in the more complex, general form of the equation. The acquisition of appropriate schemas would therefore seem to be particularly advantageous to learning in this domain because it would enable appropriate differentiation and application of formulae, a task that new students frequently find difficult. Consequently, we expect text-editing performance to reflect the different levels of schema complexity and establishing that these expected effects are present is an important preliminary step to the investigation of the main aims of the study.

Table 1

Rules and formulas used for text-editing and solution problems

				Problem No.		
Class	Rule	Formula	Section	M	S	I
A-M	Addition Rule - Mutually exclusive events.	$P(A \cup B) = P(A) + P(B)$	TE	13	10	7
			Sol	NA	5	6
A-NM	Addition Rule - Not mutually exclusive events	$P(A \cup B) =$ $P(A) + P(B) - P(A\cap B)$	TE	8	12	2
			Sol	NA	4	3
M-I	Multiplication Rule - Independent events	$P(A \cap B) = P(A) \times P(B)$	TE	14	3	5
			Sol	NA	10	9
M-NI	Multiplication Rule - Not independent events.	$P(A \cap B) = P(A) \times$ $P(B\|A)$	TE	4	11	15
			Sol	NA	8	7
Comb.	Combined rules: A-M or A-NM *with* M-I or M-NI		TE	6	1	9
			Sol	NA	1	2

Notes: Solution problems (Sol) were constructed to be isomorphic to text-editing (TE) problems; M = missing; S = sufficient; I = irrelevant; NA = problems containing missing information were not presented for solution.

The main aim was to extend the application of text-editing to the assessment of schematic knowledge of *probability theory*, an area of statistics that has traditionally proved a considerable hurdle for students from a variety of disciplines (Konold, 1995). A subsidiary aim was to test whether the high R^2 values (.80 - .90) reported in earlier studies by Low and

Over (1989, 1990) are robust to a change in methodology. Low and Over used the same irrelevant and sufficient problems in both the text-editing and solution sections of the studies. That is, participants were first asked to text-edit a set of problems containing irrelevant, sufficient, and missing information and then asked to solve a subset of those same problems (i.e., the ones containing sufficient and irrelevant information). Repeated presentation of the same problems could inflate the correlation between text-editing and solution. The second aim of the present study, therefore, was to determine the relationship between text-editing scores and solution scores when different (but isomorphic) problems were used in both sections. Observation of high R^2 values in this more stringent situation would certainly encourage serious consideration of text-editing as a way of assessing schematic knowledge in statistics.

The third aim concerned the comparison between the text-editing approach to the assessment of schematic knowledge and the more traditional problem classification approach wherein participants are asked to classify problems in terms of problem-relatedness (Schoenfeld & Herrmann, 1982). In their first experiment, Low and Over (1990) did not find any relationship between problem solution and performance on a problem-classification task, nor was there any relationship between text-editing and problem-classification. In their second experiment, where it was more clear that the problems to be classified were “similar or different in terms of underlying structure and not surface detail” (Low & Over, 1990, p. 68), a relationship between text-editing and problem-classification was found. However, the relationship between solution and classification was not investigated. Should text-editing provide incremental prediction of solution, it would constitute further evidence that text-editing is a better technique for assessing schematic knowledge than some of these traditional techniques.

Finally, the population we have chosen to draw our sample from is very diverse. It consists of university students pursuing majors ranging from science and engineering, to business and arts and hence there may be strong individual differences in exposure to training in mathematics and statistics. Although sample diversity is desirable in demonstrating generalizability of the text-editing method, these varying rates of exposure may actually serve to qualify our comparison with the studies of Low and her colleagues who have used samples less heterogeneous in experience. Therefore, tests of the main aims will also be conducted on a homogenous sub-sample of the more quantitatively inclined science and engineering students.

Method

Participants

The participants were first and second year students enrolled in a compulsory introductory statistics unit at the University of Southern Queensland, Australia. Testing time was arranged so that students had just completed a training module on probability and were preparing for an assignment on the same topic. Of the 100 students who volunteered to participate, 10 either failed to complete any of the last section of the test or did not finish in the designated time, and were therefore omitted from the analysis. The remaining 65 males and 25 females completed the study in the one hour allocated. Students ranged in age from 18 years to 44 years with a mean of 22.3 years (*SD* = 6.50). Approximately 31% of the students came from the Faculty of Engineering, 29% from Commerce, 24% from Science, and 16% from Arts.

Instruments and Procedure

Each student completed a booklet containing three sections. In addition, students were asked to provide some biographical information regarding their student status and to note some of their self-perceptions about statistical probability. The sections of the questionnaire were as follows:

Section I - Text-editing.

Section I consisted of 15 problems presented in the text-editing format used by Low and Over (1990). Students had to classify whether problems contained (a) *sufficient* information for solution, (b) *irrelevant* information that was not necessary for solution, or (c) *missing* information required to solve the problem. These problems (five for each text-editing category) were drawn from the curriculum of the introductory statistics unit in which the students were enrolled (University of Southern Queensland, 1993). All problems were vetted by teaching staff to ensure the problems covered information students were required to know for assessment. An example of a problem used from each of these categories follows:

Missing Information. You have just bought 5 tickets in a local raffle. Your brother has bought 12 tickets. What is the probability that you or your brother will win?

Sufficient Information. From an ordinary deck of 52 playing cards, one card is selected at random. What is the probability that it is a picture card (Jack, Queen, King) or a 2?

Irrelevant Information. Two cards are drawn randomly from an ordinary pack of 52 playing cards. This is done with replacement. If 25% of the cards are diamonds, what is the probability of obtaining a 9 and a Queen in that order?

To ensure wide coverage of the domain, problems were selected from five classes based on either simple or complex conjunction ($A \cup B$) and simple or complex disjunction ($A \cap B$). These classes, shown in Table 1, can be considered to be determined by the formula that would be implemented by an appropriately cued schema[2]. Care was taken to ensure that each class consisted of a sufficient, irrelevant and missing problem although respondents were not made aware of this. The problems were presented in the same randomised sequence to all students. Again, we expect that problem classes are not of equal difficulty and we will take this into consideration in the analyses.

Section II - Classification of problem relatedness.

In this section, participants classified problems in terms of similarity (i.e., same or different). Twenty problems selected from the sufficient and irrelevant problems used in Section I and Section III (described below), were presented in ten pairs. Five of the ten pairs contained problems with different structural details (or formulae required to solve them), and five pairs contained similar structural details. Each member of a pair differed in terms of surface details so as to reduce the possibility of students making a comparison solely on this basis (Low & Over, 1990). Students were told at the beginning of the session that classification of relatedness should be made in terms of the strategies and procedures necessary to solve the two problems.

Section III - Solution.

Section III required participants to solve ten problems based on those from Section I that contained sufficient and irrelevant information. Problems missing information were not used in this section because they do not contain sufficient information for solution. These problems were altered so that while they used the same structural formats as those from Section I, the surface details were different (i.e., they were developed to be isomorphic). Table 1 indicates the text-editing problem from which each solution problem was derived. Students were asked to show their working.

General Procedure

Testing was conducted in group sessions. The instructions for each section were read to the students at the start of the session so that on completion of one section students did not need to wait for others to finish. The limited testing time necessitated this procedure. Written instructions to the same effect were also provided. In addition, students were asked not to go back to a previous section once they had completed it. They were encouraged to work as quickly and accurately as possible and to ask for assistance if they had any queries.

Results

Preliminary Analysis

Scoring

Two types of text-editing scores were derived from Section I: (1) a text-editing score for each category aggregating across problem class and (2) a total text-editing score. For a problem to be scored as correct, students had to classify the problem as belonging to the correct text-editing category and also to identify the missing or irrelevant information where appropriate. The classification score in Section II was the number of correct classifications made over the ten pairs of problems. Partial-credit scoring was used for solution in Section III. A correct answer was scored "2", a partially correct answer was scored "1" (e.g., where students applied a correct formula but made minor calculation errors), and an incorrect answer was scored "0". Each score was converted to a proportion of the maximum possible for that task (text-editing; classification; solution) to facilitate comparisons across tasks. Means, standard deviations, and pairwise correlation for these measures are shown in Table 2.

Table 2

Descriptive statistics and correlations for (a) text-editing, classification, and solution, and (b) text-editing based on problems class and solution.

(a)

Variable		Mean	(sd)	(1)	(2)	(3)	(4)	(5)
TE-missing	(1)	0.35	(0.25)					
TE-irrelevant	(2)	0.24	(0.23)	.67***				
TE-sufficient	(3)	0.84	(0.22)	.30**	.24*			
TE-total	(4)	0.48	(0.18)	.86***	.83***	.64***		
Classification	(5)	0.51	(0.18)	.41***	.26*	.18	.37**	
Solution	(6)	0.34	(0.26)	.61***	.51***	.31**	.62***	.25*

(b)

Variable		Mean	(sd)	(1)	(2)	(3)	(4)	(5)
TE-class A-M	(1)	0.53	(0.26)	-				
TE-class A-NM	(2)	0.47	(0.21)	.44***	-			
TE-class M-I	(3)	0.58	(0.36)	.46***	.37***	-		
TE-class M-NI	(4)	0.47	(0.16)	.21*	.10	.16**	-	
TE-class Comb.	(5)	0.64	(0.33)	.54***	.32**	.55***	.27*	-
Solution	(6)	0.34	(0.26)	.52***	.26**	.51***	.19**	.57***

p < .05, ** *p* < .01, *** *p* < .001; N = 90; A-M = addition-mutually exclusive; A-NM = addition-not mutually exclusive; M-I = multiplication-independent; M-NI = multiplication-not independent; Comb = Combination of rules

By way of preliminary analyses, we checked for patterns among the Table 2 means to determine whether performance was in accordance with expectations of text-editing category effects and differences in problem class complexity. As expected, the main-effects for both text-editing category and problem class were significant: Pillai's Trace = .57, $F(2, 86) = 28.25$, $p < .001$; and Pillai's Trace = .83, $F(4, 88) = 219.44$, $p < .001$, respectively[3]. Univariate analyses confirmed the ordering of the text-editing categories with *irrelevant* problems significantly more difficult than *missing* problems, $F(1,178) = 29.87$, $p < .001$, and both of these more difficult than *sufficient* problems, $F(1,178) = 279.45$, $p < .001$. Regarding the problem classes, problems based on Class A-M (addition rule - mutually exclusive events) were significantly easier than problems based on Class A-NM (addition rule - events *not* mutually exclusive), $F(1, 356) = 102.47$, $p < .001$; and problems based on Class M-I (multiplication rule - independent events) were significantly easier than problems based on Class M-NI (multiplication rule - non-independent events), $F(1, 356) = 20.64$, $p < .001$. This is consistent with expectations because both A-NM and M-NI entail more complex instantiations of the basic underlying rule. Interestingly, there were no significant differences between A-M and M-I (simple addition and simple multiplication), $F(1, 356) = 2.26$, $p = .13$, nor between A-NM and M-NI (complex addition and complex multiplication), $F(1, 356) < 1$.

Continuing these preliminary analyses, we then examined performance on the 10 problems (5 containing sufficient information and 5 containing irrelevant information) that participants were asked to solve in Section III. Results indicated that there was no difference overall in the difficulty of solving problems containing sufficient or irrelevant information (Pillai's Trace = .03, $F(1, 89) = 2.28$, $p = .135$). There was however a significant main effect for problem class (Pillai's Trace = .20, $F(4, 86) = 5.26$, $p = .001$) where problems based on addition rules (A-M and A-NM) were significantly easier to solve than the other problem classes (M-I, M-NI, and Comb.).

As a final step in the preliminary analyses, we then compared the relative difficulty of text-editing and solution as a function of problem class. The solution task is based on only irrelevant and sufficient items and therefore to allow comparison the text-editing score was recalculated without the items that had missing information. The interpretation of the effect of problem class on this new text-editing score remained unchanged. Overall, it was easier to text-edit problems than to solve them (Pillai's Trace = .445, $F(1, 89) = 71.25$, $p < .001$). We have already considered the effects of problem class on text-editing and solution separately, so interpretation of these results will not be repeated. Of interest here is the difference between text-editing performance and solution as a function of problem class. The results indicate that although on the whole text editing was easier than solution, this effect was more pronounced for the multiplication and combined problems (M-I, M-NI, Comb: Pillai's Trace = .497, $F(1, 89) = 87.91$, $p < .001$) than for the addition problems (A-M, A-NM: Pillai's Trace = .056, $F(1, 89) = 5.29$, $p = .024$). In combination with the separate analyses reported above, these findings suggest that text-editing ability is not equally aligned with solution performance across problem classes.

Examining the relationship between text-editing and problem solution

The main aim of the present study was to test whether the Low and Over (1989, 1990) findings in the field of mathematics regarding the robustness of text-editing as a predictor of solution could be replicated in the field of statistics. Following their method, we used a two-step approach, focusing initially on correlations and regression analysis before moving to a detailed analysis of performance on individual text-editing items containing sufficient and irrelevant information and subsequent performance on the solution of parallel versions (isomorphs) of those same items. The correlations and regression analyses are presented first.

Pairwise correlations among the performance measures are reported in Table 2. The first three variables in Table 2A represent the three text-editing categories. The fourth variable is the total text-editing score, formed by averaging across all items. The fifth variable is the classification task that has been used in past research as a measure of schematic knowledge and was included here as a comparative measure. Table 2B contains text-editing scores based on problem class aggregated across text-editing category. It can be seen from Table 2A that whilst the three text-editing measures had moderate to strong correlations with solution, the relationship between the classification task and solution was much weaker ($r = .25$, $p < .05$). To explore the links between this set of predictor variables and solution scores, two multiple regression analyses were conducted. The first analysis explored the relationship between solution and text-editing scores based on sufficient, missing, and irrelevant problems, and the problem-classification task. Together, these variables accounted for 40.5% (37.7% adjusted) of the variability in solution scores ($R^2 = .405$, $F(4,85) = 14.489$, $p < .001$). Text-editing problems with missing information was the only measure that contributed uniquely to the prediction of solution score and accounted for nearly one quarter of the total explained variance ($sr^2 = .097$; $\beta = .447$, $t(85) = 3.72$, $p <.001$). The second regression analysis was for problem class (Table 2B) and therefore focused on a slightly different way to partition the text-editing variance. Together the five variables accounted for 42.2% (38.8% adjusted) of the variation in solutions scores, $R^2 = .422$, $F(5, 84) = 12.28$, $p < .001$. Text-editing of simple addition rule problems (class A-M) uniquely accounted for 4.0% ($sr^2 = .04$, $\beta = .26$, $t(83) = 2.40$, $p =.018$), text-editing problems based on simple multiplication rules (class M-I) accounted for 3.5% ($sr^2 = .035$, $\beta = .23$, $t(83) = 2.24$, $p =.027$), and text-editing problems based on a combination of simple addition and simple multiplication rules (class Comb) uniquely accounted for 5.6% of the variation in solution ($sr^2 = .056$, $\beta = .31$, $t(83) = 2.84$, $p = .005$). There were no other unique predictors. This tends to indicate that all else being equal, solution is best predicted by variability in text-editing problems that entail better-mastered schemas. This is an interesting result particularly when one considers the fact that the simpler rules are typically taught first, followed by instruction in the more complex forms. Hence, schemas for these simple problems are more likely to be better developed than for the more complex forms. The results suggest that the text-editing task was sensitive to variation in these simple types of problems – where knowledge is more likely to be schematic rather than declarative or procedural.

To summarise findings regarding text-editing and solution performance at the individual item level, we used the same technique reported by Low and Over (1989, 1990). Table 3 shows the number of students who correctly classified appropriate problems as containing sufficient or irrelevant information (TE+) and who were incorrect in this classification (TE–) relative to their success (+) or failure (–) on solution of the isomorphic problem in Section III. This table considers all events over the ten problems and 90 students. There were many instances where a student was able to correctly text-edit a problem but then failed to solve the structurally equivalent problem. Of the 485 instances where a student was successful in text-editing a problem, there were 334 (68.9%) incorrect attempts at solving a formally identical problem. This is consistent with previous findings of Low and her colleagues showing that text-editing is a necessary but not sufficient condition for solution.

Table 3

Frequency of correct (TE+) and incorrect (TE–) responses for (a) sufficient and (b) irrelevant text-editing problems by solution success (+) and failure (–) on isomorphic problems.

(a)

	Sufficient									
	A-M		A-NM		M-I		M-NI		Comb.	
	SOL05		SOL04		SOL10		SOL08		SOL01	
	+	–	+	–	+	–	+	–	+	–
TE+	26	55	35	43	18	43	21	59	6	73
TE–	0	9	1	11	3	26	0	10	0	11

(b)

	Irrelevant									
	A-M		A-NM		M-I		M-NI		Comb.	
	SOL06		SOL03		SOL09		SOL07		SOL02	
	+	–	+	–	+	–	+	–	+	–
TE+	11	4	4	2	17	27	0	4	13	24
TE–	34	41	22	62	6	40	27	59	5	48

Note: A-M, A-NM, M-I, M-NI, and Comb. represent problem class (see Table 1). For these comparisons a success on solution was defined as use of a correct formula and therefore a score of 1 or 2 was considered a success.

Further examination of Table 3 shows that of the 415 occurrences where students were *unable* to text-edit a problem (TE-), 98 were *able* to pass the relevant solution problem (+). In 94 of the 344 (27.3%) possible instances where a student failed to correctly text-edit an *irrelevant* problem (Table 3A), students were able to solve the parallel problem. Alternatively, in only 4 of the 71 occurrences (5.6%) where a student failed to correctly text-edit a *sufficient* problem (Table 3B), did he or she correctly solve the parallel problem in Section III. In general, these results are not predicted by the theory. However it seems clear from Table 3 that most of the students who failed text-editing and passed solution did so because they misclassified the irrelevant problem as *sufficient*. From the perspective of a novice, these errors may have been made on somewhat “permissible” grounds. Irrelevant problems do indeed have sufficient information to be solved. If these cases are removed then in only 3.8% of the total number of cases where a student failed to text-edit a problem, did a pass in Section III occur.

Group differences: A supplementary analysis.

Students were classified on the basis of whether they were pursuing a predominately quantitative degree (Science: engineering and science majors, n = 50) or not (Non-science: commerce and arts majors, n = 40).

Table 4

Performance of Science and Non-science students: Means (SD's in parentheses) and test of group differences.

	Science	Non-science	Test of Difference		
Variable	n=50	n = 40	t	df	p-value
TE-missing	0.43 (0.25)	0.25 (0.22)	3.54	88	0.001
TE-sufficient	0.84 (0.23)	0.84 (0.20)	0.09	88	0.931
TE-irrelevant	0.30 (0.23)	0.16 (0.21)	2.94	88	0.004
TE-class A-M	0.64 (0.25)	0.55 (0.23)	1.75	88	0.084
TE-class A-NM	0.36 (0.23)	0.33 (0.13)	0.65	88	0.518
TE-class M-I	0.65 (0.37)	0.43 (0.31)	3.05	88	0.003
TE-class M-NI	0.37 (0.17)	0.38 (0.17)	-0.23	88	0.818
TE-class Comb.	0.60 (0.32)	0.40 (0.23)	3.36	88	0.001
Classification	0.53 (0.17)	0.47 (0.18)	1.60	88	0.112
Solution	0.43 (0.28)	0.23 (0.19)	3.80	88	0.000

Table 4 reports the group mean scores for the measures from Table 2 and indicates, as might be expected, that Science students tended to perform significantly better than Non-science students on most measures. The focus of these supplementary analyses is on members of the Science group who are likely to be more homogenous in mathematics and statistics experience, and hence, the regression analyses were repeated for only these students. In this sub-sample, 50% of the total variation in solution scores was accounted for by text-editing category measures and problem classification (R^2 = .50, adjusted R^2 = .46; $F(4, 45) = 11.41$, $p < .001$). Although when text-editing performance was partitioned into the five problem classes (R^2 = .53, adjusted R^2 = .476; F(5, 44) = 9.89, p < .001). The general pattern of results was unchanged (TE A-M, p = .011; TE M-I, p = .007; TE Comb., p = .071).

[insert Table 4 about here]this is a slight increase in R^2 when compared to the combined sample, the pattern of results did not change. The TE-missing measure remained the only unique predictor of solution (*p* = .001). Similarly, a slight increase in the total amount of variance in solution score was accounted for

Discussion

The text-editing task rests on the assumption that if a student can specify the necessary and sufficient information in the text of a problem required to solve it, then he or she possesses domain appropriate schematic knowledge (Low & Over, 1989, 1990, 1992, 1993; Low et al., 1994; Ngu et al., 2002). Using problem solution as the benchmark for the attainment of schematic knowledge, we found support for this proposition but the strength of the association was less robust than that reported by Low and Over (1989, 1990). The literature also suggests that tasks such as the classification of problem relatedness are also effective measures of schematic knowledge. We did not find overwhelming support for this proposition and will begin by discussing possible reasons why the Section II classification task did not function as expected.

Some studies (Low & Over, 1989, 1990; Schoenfeld & Herrmann, 1982; Sweller, 1989) have found classification of problem relatedness to be a useful predictor of schematic knowledge. The present study found only a relatively weak relationship ($r = .25$, $p < .05$) that did not translate into a significant beta weight (i.e., incremental validity) when classification was used along with text-editing as a predictor of problem solution. In our view, these findings can be reconciled if the degree of complexity involved in the task and the level of expertise of the participants are taken into consideration. There is evidence to suggest that the development of knowledge and expertise in a domain does not necessarily move from declarative, to procedural, to schematic in a smooth monotonic fashion (e.g., Rittle-Johnson et al., 2001). There is also evidence to suggest that depth of knowledge does improve as knowledge becomes proceduralized, and that expertise increases with the acquisition of schemas (Chi et al., 1982). Quite involved and precise use of schematic knowledge is required to perform well in the classification task (Cooper & Sweller, 1987; Sweller, 1989). It is likely that students in this sample had not yet acquired a mature understanding across all aspects of the problem domain and that the relative level of knowledge acquisition in this sample was not as well developed as samples from other studies. The probability domain itself is known to be particularly difficult for students to master (Konold, 1995). Discovering the conditions that determine the association between classification of problem relatedness and problem solution is an area for further research.

Although the present study found a strong general association between text-editing and problem solution, the magnitude of the relationship was somewhat lower than that reported previously. Low and Over (1990) reported that detecting missing, irrelevant and sufficient information in the text-editing task accounted for up to 80 percent of the variation in ability to solve algebraic word problems. In the current study, text-editing accounted for approximately forty percent of the variation in the students' ability to solve probability problems. This is still an impressive outcome as far as the viability of text-editing is concerned but the discrepancies between the findings require some explanation. Two reasons are proposed for this. First, as previously suggested, it is possible that the high figures in the Low and Over studies were a result of an inherent association between solving and text-editing identical problems. Our study avoided this confounding by having the surface details of repeated presentations altered while the structural or schematic details were kept the same. Second, differences between algebra and probability domains in the availability of alternative solution strategies may have contributed to the lower proportion of unexplained variability. The high R^2 values (> .80) reported by Low and Over (1989; 1990) may suggest that there were only minor between-student variations in problem-solving strategies. Not only did our assessment span a more heterogeneous content, the students in this study were relative novices on at least some problem classes. Therefore the use of general problem-solving procedures (rather than schemas) is likely to have been more prominent and varied (Chi et al., 1982; Gick, 1986; Sweller et al., 1998) in our study. If a substantial proportion of students were using general problem-solving strategies to differentiate between appropriate formulas, it is not surprising that the variance in solution ability explained by text-editing was somewhat lower than in the case of the mathematics studies.

Furthermore, we argued that it is likely that there were strong individual differences in exposure to training in mathematics and statistics and that these varying rates of exposure may have acted as moderating influences. The size of the sample dictates some caution when analysing subgroups. This considered, although the amount of variance accounted for increased slightly when the more homogenous sub-sample of Science students was compared to the full sample, the R^2 values (around .50) did not approach the high values reported in previous work (Low & Over, 1989, 1990) and the pattern of variables predicting solution did not change.

In their work using algebraic word problems, Low and her associates found that virtually no-one was able to solve a problem having failed text-editing (Low & Over, 1989, 1990). This finding was replicated in the current study, although only after a response of “sufficient” was accepted as correct when irrelevant information was included. Low and Over interpreted this finding as indicating that the task of text-editing assesses a level of knowledge that is schematic in structure and nature. Such an interpretation is within the bounds of the declarative-procedural knowledge theory (Anderson, 1990; Anderson & Schunn, 2000) and the broader conceptualisation of schemas (Sweller et al., 1998). It has been repeatedly demonstrated that schematic knowledge is a sufficient but not a necessary condition for solution and that general problem solving procedures can be used to supplement domain appropriate performance (Low & Over, 1990; Rumelhart & Norman, 1981). Our findings support this earlier research and show that the text-editing technique can be applied to the field of basic probability theory in statistics.

Finally, some comment on the validity of measurement is warranted. It seems clear that that the classification of problem relatedness may not be a suitable measure for these particiapnts. The preliminary analyses showed that differences in solution rate as a function of problem class provide evidence that students' schemas for problems based on multiplication rules (M-I and M-NI) and problems based on multiple formulas (Comb.) may be less developed than are their schemas for addition-rule problems. Although text-editing performance reflected differences in simple versus complex instantiations of the addition and multiplication formula, it does not seem to be sensitive to the actual difficulties in solution. One possible reason for this divergence is that the recalculated text-editing score did not include items with missing information. It may be the case that identifying missing information is crucial to the sensitivity of the text-editing task to differences in knowledge.

The regression analyses indicate that text-editing problems based on simple addition and multiplication rules were most predictive of solution. It seems that students are yet to acquire sufficiently elaborated schemas for the more complex classes of probability problems. We have evidence to suggest that text-editing is an appropriate tool to use with introductory statistics students and, rather than replacing solution as an assessment technique, can be used to supplement both instruction (e.g., Ngu et al., 2002) and assessment.

Conclusions

It is impressive that text-editing, a technique designed to assess schematic knowledge without requiring students to actually solve problems, can explain up to forty percent of the variance in problem solution in such a diverse sample. From an educational point of view, it is equally impressive that students who were unable to text-edit a problem were almost certainly unable to then go on and solve it. The demonstration of this “necessary but not sufficient” link is a powerful reminder to students that mathematics is not purely about computational skills and that schematic knowledge is an essential prerequisite for skilled performance. Finally, the findings offer some support for the conceptualization of schematic knowledge as being a more advanced stage of knowledge-acquisition within the familiar declarative-procedural framework.

We close with some comments about the limitations of our study. Firstly, we did not collect any qualitative data so we cannot tell just how novel the text-editing experience was for these students or how that novelty may have affected their responses. Further studies could devote more effort to familiarising students with text-editing procedures before commencing the collection of data. Secondly, we learned for ourselves that constructing problems that fit neatly into the various categories is not an easy task. Like any other form of assessment, it is important to validate items to ensure that they perform as expected. This may have been an issue for the classification of problem-relatedness tasks where boundaries between probability problem classes can be indistinct. Thirdly, and we have already alluded to this limitation, studies of knowledge acquisition depend on sampling a range of levels of expertise. Our choice of relatively naïve students may limit the generalisability of our findings. These limitations, however, were all likely to attenuate the strength of the association between text-editing and actual problem solving rather than to suggest a connection that does not exist. Thus we conclude that the need for more innovative ways to assess statistical knowledge (Garfield, 1994) can be answered in part by using text-editing as an aid in learning and assessment.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406.

Anderson, J. R. (1990). *The adaptive character of thought* (Vol. 1). Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*(4), 703-719.

Anderson, J. R., & Neves, D. M. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57-84). Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 1-33). Mahwah, NJ: Erlbaum.

Benaroch, M. (2001). Declarative representation of strategic control knowledge. *International Journal of Human-Computer Studies, 55*(881-917).

Birney, D. P., & Halford, G. S. (2002). Cognitive complexity of suppositional reasoning: An application of the relational complexity metric to the knight-knave task. *Thinking and Reasoning, 8*(2), 109-134.

Braine, M. D. S. (1990). The "Natural Logic" approach to reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 133-157). Hillsdale, NJ: Lawrence Erlbaum.

Byrne, R. M. J., & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition, 62*, 1-49.

Carraher, T. N., Carraher, D., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology, 3*, 21-29.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology, 91*(4), 703-715.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-416.

Chi, M. T., Fletovich, P. J. H., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Sciences, 5*, 121-152.

Chi, M. T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-75). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, N. J., Poldrack, R. A., & Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory, 5*, 131-178.

Cook, M. (2001). Mathematics: The thinking arena for problem solving. In A. L. Costa (Ed.), *Minds: A Resource Book for Teaching Thinking*. VA: ASCD.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347-362.

de Jong, T., & Ferguson-Hessler, M. (1986). Cognitive structures of good and poor novice problem solvers in physics. *Journal of Educational Psychology, 78*(4), 279-288.

de Jong, T., & Ferguson-Hessler, M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113.

Garfield, J. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statitics Education, 2*(1), online.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21*(1 & 2), 99-120.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hegarty, M., & Koshevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684-689.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction*. Hove: Lawrence Erlbaum Associates.

Kirasic, K. C., Allen, G. L., Dobson, S. H., & Binder, K. S. (1996). Aging, cognitive resources, and declarative learning. *Psychology and Aging, 11*(4), 658-670.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statitics Education, 3*(1), online.

Littlefield, J., & Rieser, J. J. (1993). Semantic features of similarity and children's strategies for identifying relevant information in mathematical story problems. *Cognition and Instruction, 11*(2), 133-188.

Low, R., & Over, R. (1989). Detection of missing and irrelevant information within algebraic story problems. *British Journal of Educational Psychology, 59*, 296-305.

Low, R., & Over, R. (1990). Text editing of algebraic word problems. *Australian Journal of Psychology, 42*(1), 63-73.

Low, R., & Over, R. (1992). Hierarchical ordering of schematic knowledge relating to area-of-rectangle problems. *Journal of Educational Psychology, 84*(1), 62-69.

Low, R., & Over, R. (1993). Gender differences in solution of algebraic word problems containing irrelevant information. *Journal of Educational Psychology, 85*(2), 331-339.

Low, R., Over, R., Doolan, L., & Michell, S. (1994). Solution of algebraic word problems following training in identifying necessary and sufficient information within problems. *American Journal of Psychology, 107*(3), 423-439.

Mayer, R. E. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74*(2), 199-216.

Ngu, B. H., Low, R., & Sweller, J. (2002). Text editing in chemistry instruction. *Instructional Science, 30*, 379-402.

Novick, L. R., Hurley, S. M., & Francis, M. (1999). Evidence for abstract, schematic knowledge of three spatial diagram representations. *Memory and Cognition, 27*(2), 288-308.

Nuñes, T. (1994). Street intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. 2, pp. 1045-1049). New York: Macmillan.

Piaget, J. (1950). *The psychology of intelligence* (M. Piercy & Berlyne., Trans.): Routledge & Kegan Paul.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorise statistics word problems. *Journal of Educational Psychology, 88*(1), 144-161.

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology, 16*(3), 325-342.

Rips, L. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90*(1), 38-71.

Rips, L. (1989). The psychology of knights and knaves. *Cognition, 31*, 85-116.

Rips, L. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, Mass: MIT Press.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346-362.

Rumelhart, D. E., & Norman, D. A. (1981). Analogical processes in learning. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 335-359). Hillsdale, NJ: Lawrence Erlbaum Associates.

Savelsbergh, E. R., de Jong, T., & Ferguson-Hessler, M. (2002). Situational knowledge in physics: The case of electrodynamics. *Journal of Research in Science Teaching, 39*(10), 928–951.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(5), 484-494.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*(4), 457-466.

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology, 45*(1), 1-8.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176-192.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(5), 463-474.

Sweller, J., Mawer, R., & Howe, W. (1982). Consequences of history-cued and means-ends strategies in problem solving. *American Journal of Psychology, 95,* 455-483.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

University of Southern Queensland. (1993). *Unit 75002 data analysis study book.* Toowoomba: USQ: Distance Education Centre.

Van Gog, T., Paas, F., & Van Merriënboer, J. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32*(83-98).

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*(1), 1-30.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*(3), 319-331.

Woltz, D. J., Gardner, M. K., & Gyll, S. P. (2000). The role of attention processes in near transfer of cognitive skills. *Learning and Individual Differences, 12,* 209-251.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*(3), 137-166.

Footnotes

1. The generality of these strategies is evident in that competing theories of human reasoning typically do not dispute their importance. They are present in rule-based theories of reasoning (Braine, 1990; Rips, 1983, 1994), in theories based on mental models or representations (Halford, 1993; Johnson-Laird & Byrne, 1991), and have also been conceptualized in tasks that entail both rules and mental models (e.g., Birney & Halford, 2002).

2. We acknowledge that there are often multiple ways to solve probability problems. The categories chosen reflect the framework in which probability was taught to the participants.

3. Pillai's Trace is a multivariate test of mean differences between repeated-measures of the same individual. Multivariate tests do not require the sphericity assumption of univariate tests. Hence, we use multivariate analyses to test the omnibus effects and univariate analyses for follow-up multiple-comparisons.