Mathematics Education Research Journal 2003, Vol. 15, No. 3, 270-287

The Role of the Sampling Distribution in Understanding Statistical Inference

Kay Lipson
Swinburne University of Technology

Many statistics educators believe that few students develop the level of conceptual understanding essential for them to apply correctly the statistical techniques at their disposal and to interpret their outcomes appropriately. It is also commonly believed that the sampling distribution plays an important role in developing this understanding. This study clarifies the role of the sampling distribution in student understanding of statistical inference, and makes recommendations concerning the content and conduct of teaching and learning strategies in this area.

Over recent years there has been an expansion in the teaching of statistics at all educational levels. This increase has been due, in part, to the recognition of the importance of quantitative literacy but also to the availability of computer-based technology that can be used to carry out sophisticated statistical analyses. Many feel, however, that although increasing numbers of students study statistics, the number who gain a real appreciation of the power and purpose of statistics is extremely small (see, e.g., Garfield, 2002; Konold, 1991; Williams, 1998). To date there is little empirical evidence that increasingly refined technological support is doing much to change this (Mills, 2002). More research is needed in order to improve the structure and teaching of introductory statistics.

The focus of this research was the statistical concepts that are critical to an understanding of statistical inference, in particular the teaching and learning of the sampling distribution. Clearly, this concept is of fundamental statistical importance and many statistics educators (e.g., Rubin, Bruce, & Tenney, 1990; Shaughnessy, 1992; Tversky & Kahneman, 1971) have suggested that the sampling distribution is a core idea in the understanding of statistical inference, something that many teachers of the subject have intuitively recognised. In addition, many current students of statistics are not mathematically trained, and hence the more abstract concepts in statistics tend to be demonstrated rather than derived. Current computer technology allows ideas such as the sampling distribution to be demonstrated easily using readily available software. One need only look at the proliferation of computer activities dedicated to the Central Limit Theorem to confirm this (e.g., Finzer & Erickson, 1998; Kader, 1990; Kreiger & Pinter-Lucke, 1992; Stirling, 2002).

The aim of this research was to produce empirical evidence to determine if the educational emphasis on the sampling distribution holds the potential to enhance student understanding of statistical inference.

Theoretical Framework

The study was concerned with the relationship between students' knowledge of the sampling distribution, and the level of understanding that they demonstrated concerning statistical inference. In order to examine this it was necessary to consider what constituted knowledge in the content domain of sampling distribution, and how this knowledge could be determined and evaluated. It was also necessary to propose a model for understanding, and determine how understanding of statistical inference would be measured.

Procedural and Conceptual Understanding, and Schemas

It has been long recognised by many educators and researchers that often students are able to complete problems in statistics successfully but, at the same time, demonstrate no real understanding of the concepts inherent in these tasks (Garfield, 2002; Garfield & Ahlgren, 1988; Williams, 1998). In investigating this situation, particularly in relation to statistical inference, it is useful to differentiate between *procedural* and *conceptual* understanding. Procedural understanding describes a student's ability to carry out routine tasks successfully, whereas conceptual understanding implies that the student understands what is being done and why.

In order to think about levels of student understanding it helps to adopt a representation for the structure of knowledge. In this study learning was viewed from the constructivist position, where students are not regarded as passive receivers of information but rather as active constructors of highly personal mental structures called schema (see, e.g., Howard, 1983; Piaget, 1970). Marshall (1995) writes:

> A distinctive feature of a schema is that when one piece of information associated with it is retrieved from memory, other pieces of information connected to the same schema are also activated and available for mental processing. (p. vii)

It makes sense, then, to think of the schema as a connected network of concepts. This model fits well with the role schemas play in the construction of knowledge.

Hiebert and Carpenter (1992) suggest a useful relationship between the form of the cognitive structure and the level of understanding that is evidenced by the student. They suggest:

> Conceptual knowledge is equated with connected networks … A unit of conceptual knowledge is not stored as an isolated piece of information; it is conceptual knowledge only if it is part of a network. On the other hand, we define procedural knowledge as a sequence of actions. The minimal connections needed to create internal representations of a procedure are connections between succeeding actions in the procedure. (p. 78)

This relationship between level of understanding and the form and complexity of a student's schema gives a theoretical justification for evaluation of the student's understanding based on an analysis of the relevant schema. In order to undertake this analysis, however, it is necessary to propose a model of a schema that could be considered to represent conceptual understanding in the relevant content domain, in this case statistical inference, including the sub-domain of interest, the sampling distribution.

Analysis of the Content Domain using Concept Maps

Because it is impossible to observe an individual's schema directly, hypotheses are required about the structure of schemas appropriate for particular statistical tasks. It is necessary to look carefully at the content of the task in order to ascertain the knowledge required to carry out that task successfully. To investigate the important concepts in statistical inference an analysis of the underlying knowledge and the way in which each of the component ideas relates to the others is crucial.

This enables identification of the desirable features of a schema that will support both procedural and conceptual understanding in statistical inference.

A useful tool for doing this analysis is the concept map, a technique developed by Novak and Gowin (1984) and used for the purpose of content analysis by some educators (e.g., Jonassen, Beissner, & Yacci, 1993; Starr & Krajcik, 1990). Concept maps constitute a method for externalising the knowledge structure in a particular content domain. They are two-dimensional diagrams in which relationships among concepts are made explicit. When two or more concepts are linked together with a label then this forms a *proposition*, the formation of which is taken to indicate recognition of that aspect of the concept. According to Novak (1990), "the meaning of any concept for a person would be represented by all of the propositional linkages the person could construct for that concept" (p. 29).

Constructing a concept map requires one to identify important concepts concerned with the topic, rank these hierarchically, order them logically, and recognise relationships where they occur. In this research the concept map was used to analyse the content domain of statistical inference, making explicit the concepts and relationships between concepts that are fundamental to developing understanding in this topic, in particular sampling distribution.

In order to study and evaluate the students' schemas an external representation of that mental structure was necessary. Once again, the concept map provides a method for obtaining external representations of an individual's schema. Concept maps have been used in educational research to facilitate the study of the students' schemas before and after instruction (Novak, 1990), and as an assessment tool to give insight into students' understanding (Schau & Mattern, 1997; Shavelson, 1993). By directing students to construct concept maps for various components of the course the researchers could gain insight into the relevant student schemas.

Measuring Procedural and Conceptual Understanding

Since this research was concerned with the development of procedural and conceptual understanding in introductory statistical inference, it was necessary to determine a measure of conceptual understanding and a measure of procedural understanding for each participant. Since procedural understanding is a common focus of many tasks assessing statistical understanding, there existed a variety of tasks that could be used to measure procedural understanding. For conceptual understanding, however, few tasks were available that had been trialed and validated, and that covered the content domain under investigation. For this study such instruments needed to be developed.

From an analysis of what it means to know and understand mathematics Putnam, Lampert, and Peterson (1990) proposed that there are five key themes underpinning mathematical understanding. These are: *Understanding as Representation, Understanding as Knowledge Structure, Understanding as Connections between Types of Knowledge, Understanding as the Active Construction of Knowledge,* and *Understanding as Situated Cognition*. These themes were taken by Nitko and Lane (1990) and related to the development and measurement of understanding in statistics. Their framework was further developed by the researcher (Lipson, 2000) to develop a range of tasks to assess aspects of either procedural or conceptual understanding in the particular knowledge domain of introductory statistical inference, as shown in Table 1.

Table 1
Framework for Developing Tasks to Measure Understanding of Statistical Inference

Key theme	Related assessment tasks
Understanding as Representation	Tasks involve application of standard notation, representation, and algorithms to solve statistical problems. This would include standard applications of the *t*-test or chi-square test for example.
Understanding as Knowledge Structure	Tasks give insight into the knowledge structures of students; that is, tasks demonstrate that the student has made a connection between concepts, as demonstrated by hypothesis testing and use of confidence intervals.
Understanding as Connections between Types of Knowledge	Tasks require students to integrate formal knowledge with informal knowledge developed outside the class. This would include tasks requiring the interpretation of statistical concepts.
Understanding as the Active Construction of Knowledge	Tasks enable the teacher to monitor the development of knowledge over time, such as the creation of concept maps.
Understanding as Situated Cognition	Tasks require students to apply their knowledge in a variety of contexts, different from those previously seen and discussed in the classroom.

In this study, tasks that were developed within the classification of Understanding as Representation were considered to measure procedural understanding, as they refer to applications of standard procedures. Tasks that were developed under the other four key themes of understanding were considered to contribute to the measurement of conceptual understanding, as they required the students to have an holistic view of the processes that underlie statistical inference, their purpose, and relationships. Using this framework as a guide, known work on assessment at this time was expanded and supplemented by the researcher to give a set of tasks that covered the suggested range of aspects of understanding over the full content domain.

Research Hypotheses

This study was concerned with the relationship between students' knowledge of the sampling distribution, evidenced by analysis of their concept maps, and the levels of procedural and conceptual understanding that they demonstrated concerning introductory statistical inference, measured by the tasks developed using the framework in Table 1. The research hypotheses can thus be stated as follows:

1. Those students whose schema for sampling distribution demonstrated links to the sampling process and whose schema for statistical inference included links to the sampling distribution, would demonstrate the highest levels of conceptual understanding of introductory statistical inference.

2. The level of procedural understanding demonstrated by students would not necessarily be related to the content and form of their schemas for sampling distribution and statistical inference.

Method

The Setting of the Study

The study was conducted at an Australian university. The 23 mature-age students were undertaking graduate studies in either Social or Health Statistics. They were generally graduates from other disciplines, such as Business or Nursing, who had determined that knowledge of statistics would be helpful for them in their future careers. The study took place during the conduct of a subject called *An Introduction to Statistics* which was taught over a 13 week semester, and classes were held one evening a week for 3.5 hours. All data for the study were collected during the final 6 weeks of the course, and in the examination held one week after the end of the course.

Content Domain Analysis

In order to provide a benchmark for the evaluation of the student concept maps, the researcher and a colleague—both content experts in the area of introductory statistical inference—constructed a series of concept maps, for the sampling distribution, hypothesis testing, estimation, and statistical inference. These maps were first constructed by each of the experts individually and then, by a process of negotiation, common maps were agreed upon. These were termed the *expert* maps. From these expert maps, certain key propositions could be identified, which summarised both the knowledge domain and the connections between aspects of knowledge, and which characterised a connected schema. The expert concept map for the sampling distribution is shown in Figure 1 and the propositions identified from this are given in Table 2. Propositions are identified from the concept map by taking each pair of concepts together with the linking words and forming a statement.

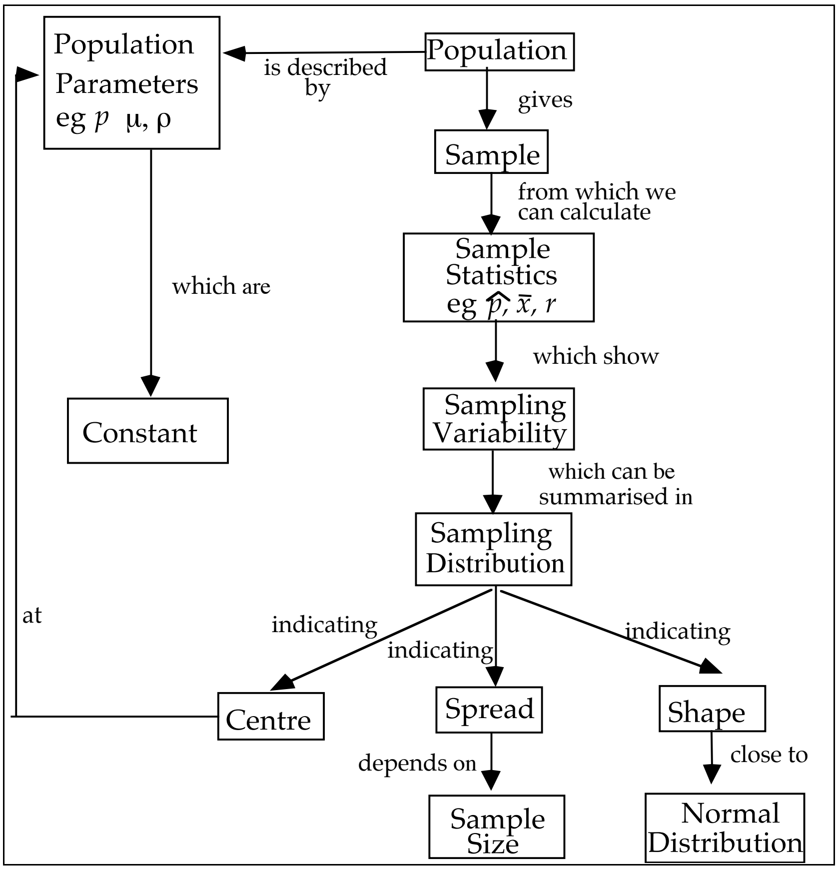

Figure 1. Expert concept map for sampling distribution.

Table 2
Key Propositions Identified in the Expert Concept Map for Sampling Distribution

Propositions
Samples are selected from populations.
Populations (distributions) are described by parameters.
Parameters are constant in value.
Samples are described by statistics.
Statistics are variable quantities.
The distribution of a sample statistics is known as a sampling distribution.
The sampling distribution of the sample statistic is approximately normal.
The sampling distribution of the sample statistic is characterised by shape, centre, spread.
The spread of the sampling distribution is related to the sample size.
The sampling distribution is centred at the population parameter.

Analysis of Student Schema Development

Previous research (Jonasson, Beissner, & Yacci, 1993) has shown that, as students learn, the schema they create becomes closer in structure to those of their instructors, and thus that the students' knowledge structure can be evaluated by comparing the students' concept maps with the expert maps. During this study the students were asked to prepare concept maps for the following topics, approximately one map each week until the final session of the course.

1. The sampling distribution of the sample proportion.
2. The sampling distribution of the sample mean.
3. The sampling distribution (general).
4. Hypothesis testing.
5. Estimation.
6. Statistical inference.

In order to facilitate the mapping process, the students were provided with a list of the key terms that had been derived from the expert maps, and then the students were asked to use these terms in the constructions of their own maps. Students were advised that the terms listed were only suggested, and any could be omitted or others added as required. The list of terms used as a starting point for each of these maps is given in Appendix 1. The purpose of the concept mapping exercises was to document the students' schemas at particular points in time. The author could identify from the maps the propositions formed by relating the terms given and then evaluate the student maps by comparison with the expert maps. This comparison was carried out not only in terms of both the number and type of propositions present, but also in terms of the links between various propositions. From a qualitative analysis of the propositions evidenced by the series of six concept maps prepared by the students over the period of the study students were categorised into groups by the relative closeness of their association to the expert maps described in the previous section and by change over time. Of importance was not merely the number of propositions included, but which ones they were. More details about this process have been reported elsewhere (Lipson, 2002).

Measures of Procedural and Conceptual Understanding

Several tasks were used to measure procedural and conceptual understanding. Some were based on the *Statistical Reasoning Assessment* instrument developed to assess conceptual understanding in probability and statistics by Konold and Garfield (1993). This series of multiple choice questions built on earlier work of Konold and others (Falk, 1993; Kahneman & Tversky, 1972; Konold, 1991). The tasks used in the study were designed to measure procedural and conceptual understanding in statistical inference, over the relevant content domain. To ensure all content areas were covered, there were seven tasks pertaining to the measurement of procedural understanding, each one concerned with a different sub-section of the content domain. Example 1, shown in Figure 2, a routine problem concerned with the measuring the students' ability to carry out a standard *t*-test from first principles, is an example of such a task.

Four tasks were used to measure conceptual understanding. Examples of three of these tasks, classified according to the framework of Nitko and Lane (1990), are given in Examples 2, 3, and 4 in Figures 2 and 3. Adapted from the *Statistical Reasoning Test* (Konold & Garfield, 1993), based on earlier work of Kahneman and

Tversky (1972), Example 2 was classified as Understanding as Connections between Types of Knowledge. Tversky and Kahneman (1972) found that 56% of undergraduate students incorrectly gave the answer C, suggesting that the majority of students believe that the variability of the sampling distribution is independent of the sample size. Obtaining the correct answer, B, implies that the student appreciates that the variability in the sampling distribution of the sample proportion is larger when the sample size is smaller. Response A indicates that the variability of the sampling distribution of the sample proportion is seen to increase with the sample size.

Example 1 (Procedural knowledge)

According to a Census held in 1956, the mean number of residents per household in an inner suburb, Richthorn, was 3.6. In 1995, a student randomly sampled 11 households from the suburb and recorded the number of residents in each with the following results

2	2	5	1	1	3	4	2	4	3	1

Can the student conclude that the mean number of residents per household in Richthorn has decreased since the 1956 Census?

Example 2 (Conceptual knowledge)

Half of all newborn babies are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day which hospital is more likely to records 80% or more female births?

A Hospital A (with 50 births a day).

B Hospital B (with 10 births a day).

C The two hospitals are equally likely to record such an event.

Example 3 (Conceptual knowledge)

A radio station claims to its advertisers that 20% of 18–25 year olds listen to this station between 6.00 pm and mid-night on weeknights. A market research company carries out independent research on behalf of an advertiser and finds that only 15% of their sample of 18–25 year olds listen to the radio station in this time period. The advertiser concludes that the radio station is misleading them. What do you think? Try to include all the relevant reasons for your answer.

Figure 2. Examples of tasks used to assess procedural and conceptual understanding (see also Figure 3).

Example 3 (Figure 2) was classified as Understanding as Situated Cognition and was designed to ascertain which of the relevant statistical schema are activated when the students are asked to consider a real world context. The question was quite intentionally open ended, and contained insufficient information for an exact answer to be obtained using a standard algorithm.

Example 4 (Figure 3) was designed to elucidate further the students' conceptual understanding of statistical inference, by requiring them to interpret the (procedural) steps in the hypothesis test in their own language. By linking the formal notation and algorithms with which they were familiar with informal knowledge that can be understood by most people without specialist statistical training, students could demonstrate Understanding as Connections between Types of Knowledge, an aspect of conceptual understanding.

Example 4 (Conceptual knowledge)

As a part of your research, you are investigating the relationship between the intelligence of a child and the intelligence of his or her mother. To this end, you administer intelligence tests to the mother and eldest child of a randomly selected sample of 30 families. A scatterplot of the data obtained indicated the presence of a moderate linear relationship between the intelligence test score of the mothers and their children and the value of Pearson's r was found to be r = 0.5135. You carry out a hypothesis test as shown below and conclude that the data you have supports your long held contention that there is a relationship between the intelligence of children and the intelligence of their mother in the general population.

A friend, who is very interested in your research, but who understands little about statistics, asks you to explain to him how you have come to this conclusion. He is happy that there is a relationship between the mother's and the eldest child's intelligence in the sample, but can't understand how you can generalise your result to include mothers and eldest children in general. You explain that this is the purpose of the hypothesis test you performed.

In the space provided in the table give a brief explanation of each step in the hypothesis test that you have carried out, so that your statistically illiterate friend is able to understand what you have done and how you were able to draw your conclusion. You can assume that your sample is properly representative of the general population.

Steps in your hypothesis test	Explanation
Hypotheses: H_0: $\rho = 0$, H_1: $\rho \neq 0$	
Significance level: $\alpha = 0.05$	
Test statistic: r = 0.5135 for n = 30 pairs of data values	
P-value: P-value = 2 x P(r > 0.5) = 2 x 0.0025 = 0.005 or 0.5%	
Decision & conclusion: As p< 0.05, reject H_0 and conclude that there is a relationship between the intelligence of children and their mothers in the general population.	

Figure 3. Additional example of a task used to assess conceptual understanding (see also Figure 2).

A compete description of all tasks used, their rationale, and interpretation, can be found in Lipson (2000). In order to determine if the tasks developed did actually measure the separate underlying constructs of procedural and conceptual understanding, factor analysis was used. An exploratory factor analysis using principle components extraction and an oblimin rotation resulted in resolution into two factors. The rotated factor loadings with variables sorted into factors and factor loadings less than 0.1 suppressed is shown in Table 3.

The factor matrix showed simple structure, and the resultant factors were clearly identifiable as measures of procedural understanding (Factor 1) and conceptual understanding (Factor 2). The correlation between Factor 1 and Factor 2 was 0.254, indicating a weak positive relationship between the two factors, due perhaps to a general underlying ability factor. This analysis provided empirical validation for the measurement process used in the study.

Table 3
Rotated Factor Loadings for Tasks Used in the Study

Items	Factor 1 Procedural	Factor 2 Conceptual
Procedural 1 (Example 1)	0.981	
Procedural 2	0.944	
Procedural 3	0.936	-0.218
Procedural 4	0.885	
Procedural 5	0.833	
Procedural 6	0.797	0.188
Procedural 7	0.718	
Conceptual 1 (Example 4)		0.828
Conceptual 2 (Example 3)		0.776
Conceptual 3 (Example 2)	-0.111	0.555
Conceptual 4	0.221	0.524

Note: Factor loadings less than 0.1 have been suppressed.

At the end of the course the students completed these procedural and conceptual tasks as a component of their final examination. At this time they also constructed the final of the six concept maps, that for *statistical inference*. Factor scores were generated for each student on the measures procedural understanding and conceptual understanding, using the coefficients obtained from the factor analysis.

In order to address the research hypotheses the researcher documented and evaluated the schema development of each student with regard to sampling distribution and the links the student constructed and maintained between sampling distribution and statistical inference. This enabled the students to be divided into groups on the basis of their knowledge of sampling distribution. The mean scores for conceptual understanding and the mean scores for procedural understanding were then compared across the groups, using analysis of variance.

Results

Classification of Students by Schema Development

The concept maps showed that for some students the sampling distribution has no place in statistical inference, and was absent entirely from their concept maps, whereas other students included the sampling distribution but did not link it to inference in a meaningful way. From the analysis of the set of six concept maps prepared by each student it was possible to document the conceptual development of the student in the domain of introductory inference over a period of time. On this basis each student was allocated into one of three broad categories. These were:

- Group 1: Students whose schemas showed evidence of the development of the concept of sampling distribution and its integration into their schema for statistical inference (7 students).
- Group 2: Students whose schemas showed evidence of the development of the concept of sampling distribution but did not integrate this into their schema for statistical inference (10 students).
- Group 3: Students who did not show evidence of the development of the concept of sampling distribution (6 students).

The concept maps for sampling distribution (Map 3 in the series of maps completed) for one student from Group 1, and for one student from Group 3, are shown in Figure 4.

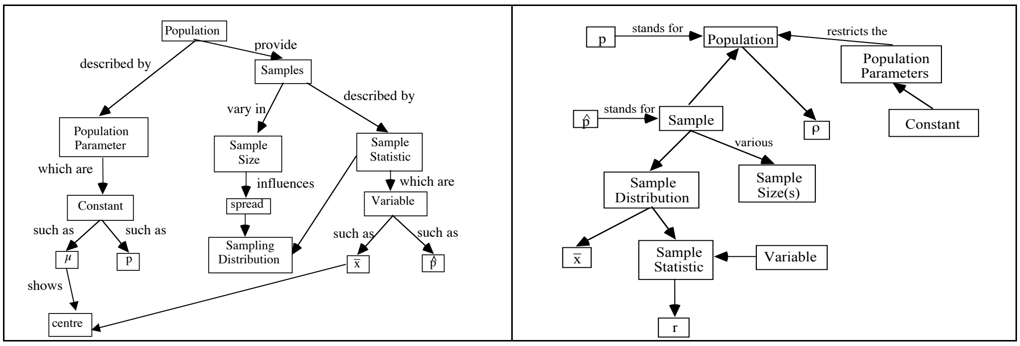

Figure 4. Concept maps for sampling distribution for one Group 1 student (left) and one Group 3 student (right).

The concept map prepared here by the Group 1 student is clearly structured showing the sampling distribution correctly as describing the sample statistic, and with the influence of sample size on the spread of the distribution clearly noted. This map included six of the ten propositions identified from the expert map. From the concept map prepared by the Group 3 student it can be seen that, although this map included four of the ten propositions identified in the expert map and the term *sample distribution* is used (correctly) to describe the distribution of the sample, the term *sampling distribution* has not been included, even though this was given as the subject of the map. Thus there is no evidence here for the formation of a schema in which sampling distribution is recognised as the distribution of the sample statistic.

The Relationship between Students' Conceptual Structures and Understanding

The means and standard deviations of the factor scores for each of the three groups are given in Table 4. The factor scores are standardised so that each variable has a mean of 0 and a standard deviation of 1. The data in Table 4 enable comparisons to be made among the three groups for the measures of procedural and conceptual understanding. For conceptual understanding it can be seen that the mean score for students in Group 1 ($M = 0.838$) was higher than that for Group 2 ($M = -0.007$), which was in turn higher than that for Group 3 ($M = -0.966$). The variation in scores was small for the Group 1 students compared to Group 2 and 3 students, although there was no statistically significant difference between the variances, which is to be expected with such small sample sizes. A one-way Analysis of Variance revealed that there was a significant difference in the scores for conceptual understanding between the three groups ($F(2,20) = 9.142$, $p = 0.002$, $\eta^2 = 0.477$). Further comparison of the mean scores showed that, as suggested in hypothesis 1, the mean score for conceptual understanding in Group 1 was significantly higher than that for Group 2 ($t(20) = 2.259$, $p = 0.018$), and that the mean score for conceptual understanding in Group 2 was significantly higher than that for Group 3 ($t(20) = 2.451$, $p = 0.012$).

Table 4
Summary statistics for factor scores for Procedural and Conceptual Understanding for each group

Group			Procedural understanding		Conceptual understanding	
		N	*M*	*SD*	*M*	*SD*
1	Sampling distribution concept developed and linked to inference	7	0.384	0.205	0.838	0.968
2	Sampling distribution concept developed but not linked to inference	10	0.004	1.010	-0.007	0.589
3	Sampling distribution concept not well developed	6	-0.455	1.311	-0.966	0.741

From Table 4 it can also be seen that with regard to procedural understanding, in this sample Group 1 students scored higher on average ($M = 0.384$) than students in Group 2 ($M = 0.004$), who in turn scored higher than students in Group 3 ($M = -0.455$), but these differences are quite small. As suggested in hypothesis 2, there was no significant difference in the mean scores for procedural understanding between the groups, ($F(2,20) = 1.154$, $p = 0.336$). Given the size of the study, however, it is not possible to reach a conclusion about the relationship between the role of the sampling distribution in the student's conceptual structure and the level of procedural understanding shown. It appears, though, that the relationship between the student's conceptual structure and the level of conceptual understanding is stronger than the relationship between the student's conceptual structure and level of procedural understanding, at least in this group of students.

The weak positive correlation ($r = 0.254$) between the measures of procedural understanding and conceptual understanding for these students also confirms that high scores in conceptual understanding are not necessarily associated with high scores in procedural understanding, and vice-versa. For example, the student with the highest conceptual understanding score ranked only eleventh on procedural understanding, whilst the student who ranked twenty-second on conceptual understanding ranked seventh on procedural understanding.

In conclusion, the research hypothesis concerning the role of the sampling distribution in the students' conceptual structure and the level of conceptual understanding exhibited, has been supported by these analyses, to some extent at least. The results have shown that there is a statistically significant difference in the levels of conceptual understanding demonstrated by the students when they are categorised into groups on the basis of the knowledge structure they exhibit regarding sampling distribution. An effect size of 0.477 indicates that these differences can be considered as quite large.

Discussion

This study has provided some empirical evidence supporting the long-held contention that knowledge of the sampling distribution is associated with the development of conceptual understanding in statistical inference. It has also confirmed the belief, held by many statistics educators, that conceptual understanding does not necessarily develop in the same way as procedural understanding in statistical inference.

There are a number of implications for the teaching of statistics that arise from this research. In particular the study has implications for the teaching of the sampling distribution, and for the assessment of students' understanding of introductory inference.

Developing Understanding of the Sampling Distribution

This research has shown that knowledge of the sampling distribution and integration of this knowledge with the concepts and practise of statistical inference as measured by an analysis of concept maps is associated with higher levels of conceptual understanding of introductory statistical inference as measured by specific problem-based tasks. Typically, introductory courses in statistics include a study of descriptive statistics, introducing methods of describing, displaying, and summarising realisations of a variable based on a sample (e.g., Moore & McCabe, 1999). The distribution of the variable in the population is also usually discussed, generally in relation to a probability distribution. For many students, and their teachers, the distribution of the population is seen as the 'true' distribution whereas the sample distribution is seen as a particular case of the theoretical population distribution. Often in texts dealing with empirical distributions, the implication is made that if there were enough data, plotted on a histogram with small enough intervals, the probability distribution function would be obtained.

An equally valid and necessary representation of the distribution of the sample statistic, however, is the empirical representation of the sampling distribution, based on repeated sampling. This is the representation which is formed when students participate in computer simulation activities, and is readily related to the process of sampling. To make the link between the empirical and theoretical representations, it can be demonstrated that, under certain assumptions, the empirical sampling distribution can be *modelled* by a known probability distribution, and that this known distribution can be used in the determination of P-values. The theoretical content analysis underpinning this study identified that the establishment and maintenance of the link between the empirical and theoretical representations of the sampling distribution was an important feature of any teaching/learning strategy. This study has confirmed that in order to develop conceptual understanding of the procedures of statistical inference the empirical representation of the sampling distribution is an important component of the student's schema for sampling distribution. Thus, teaching and learning strategies in introductory statistical inference should include the necessary pedagogic actions to facilitate the development and maintenance of this link in the students' mental structures. For example, some computer simulations (e.g., Stirling, 2002) allow the theoretical sampling distribution to be superimposed on the empirical sampling distribution so that the link between the two representations of the sampling distribution can be made explicit.

Implications for Assessment in Statistics

If the goal of the educator is to assess both conceptual and procedural understanding in their students, then traditional application questions are not sufficient. The weak association between the measures of procedural understanding and conceptual understanding for these students confirms the need to include both types of assessment tasks when evaluating student learning. Had only the traditional skills-based measurement tasks such *t*-tests or chi-square tests been used, all but two students would have been deemed to have performed at a 'satisfactory' level (that is, passed the examination). An analysis of tasks designed to elicit deeper understanding, however, revealed substantial variation in the level of conceptual understanding demonstrated.

At the time this study began, the statistics education profession had not addressed the issue of assessment of understanding in statistics. Some suggestions had been made concerning assessment, but the tasks suggested were developed in isolation and in an ad hoc way, which made it difficult to relate these tasks to a measure of student performance in a subject. An additional consideration in assessing larger classes is ensuring that assessment tasks can be used with students who have only limited contact with the instructor, often in groups. For these students tasks that take a lot of time, or involve intense one-to-one interaction such as in interviews, are really not feasible.

An analysis of the literature concerned with assessment in statistics shows that, whereas assessment in statistics has become of greater interest to researchers in recent years, there is still a lack of instruments with which to assess conceptual understanding in statistical inference. However, many valuable suggestions have been made concerning the possible style and form of assessments that could be used (e.g., Gal & Garfield, 1997). The tasks developed for measuring conceptual understanding in this study are consistent with current recommendations from the profession, and these tasks could be valuable to other researchers and educators for inclusion with or without modification into their assessment programs.

It is worth noting that whatever assessment tasks are used, it is necessary for the educator to remain aware of the students' active preference for procedural learning, and their consequent tendency to "practice" even novel questions until a procedure is created. Hubbard (1997) points out that

> If an instructor produces a non-standard question and keeps repeating it, then it becomes a standard question and students will learn a standard response. (para. 13)

In order to reveal students' levels of conceptual understanding—and to encourage them to develop such understanding—families of tasks need not only to be developed but also to be modified continually so as not to become proceduralised.

Conclusion

It should be noted that the tasks developed for the purpose of assessing procedural and conceptual understanding here represent only a starting point in a vast and challenging mission. Further thought needs to be given to the nature of tasks that could be used to measure understanding in statistics, particularly those that have application in the classroom.

The findings of this study have emanated from a small, and in some sense specialised, group of students. Although the research hypotheses have been supported within this group, the generalisability of the results to other students, particularly undergraduates, is not theoretically justifiable without replication of the research with other student groups of more diverse age and academic backgrounds.

The data collected and analysed here, however, do support the conclusion that for some students both procedural and conceptual understanding of statistical inference did develop over the period of this study. For other students, however, there was little evidence of conceptual understanding. Was this the result the students' preconceptions before the course, of the instructional strategy used, of the students' attitude to their learning, or some other factors that have not been considered? These issues were not directly addressed, and they need to be included in future research in order to understand more fully the nature of the process by which students acquire knowledge in statistical inference.

Acknowledgements

The author wishes to acknowledge the valuable contribution made by Professor Peter Jones to both the conduct of the study and the preparation of this paper.

References

Cohen, S., & Chechile, R. A. (1997). Probability distributions. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 253-262). Amsterdam: IOS Press and the International Statistics Institute.

Cohen, S., Smith, G. E., Chechile, R. A., & Cook, R. (1994). Designing software for conceptualizing statistics. In L. Brunelli & G. Cicchitelli (Eds.), *Proceedings of the first scientific meeting of the International Association for Statistical Education* (pp. 237-245). Perugia, Italy: University of Perugia.

Falk, R. (1993). *Understanding probability and statistics: A book of problems*. Wellesley, MA: A. K. Peters.

Finzer, W. F., & Erickson, T. E. (1998). *Dataspace* - A computer learning environment for data analysis and statistics based on dynamic dragging, visualisation, simulation and networked collaboration. In L. Pereira-Mendoza (Ed.), *Proceedings of the 5th International Conference on Teaching Statistics* (Vol. 2, pp. 825-829). Voorburg, The Netherlands: International Statistics Institute.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-16). Amsterdam: IOS Press and the International Statistics Institute.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3). Retrieved 28 January 2004 from the World Wide Web: http://www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*, 44-63.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York: Macmillan.

Howard, D. V. (1983). *Cognitive psychology*. New York: Macmillan.

Hubbard, R. (1997). Assessment and the process of learning statistics. *Journal of Statistics Education, 5*(1). Retrieved 28 January 2004 from the World Wide Web: http://www.amstat.org/publications/jse/v5n1/hubbard.html

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Kader, G. (1990). Simulations in mathematics: Probability and computing. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics* (Vol. 1, pp. 178–186). Voorburg, The Netherlands: International Statistics Institute.

Kahneman, D., & Tversky, A. (1972). Subjective probability. A judgement of representativeness. *Cognitive Psychology, 3*, 430-453.

Konold, C. (1991). Understanding student's beliefs about probability. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139-156). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Konold, C., & Garfield, J. (1993). *Statistical reasoning assessment. Part 1: Intuitive thinking.* Unpublished manuscript, Scientific Reasoning Research Institute, University of Massachusetts, Amherst.

Kreiger, H., & Pinter-Lucke, J. (1992). Computer graphics and simulations in teaching statistics. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century. MAA Notes No 26* (pp. 198-206). Washington: Mathematical Association of America.

Lipson, K. (2000). *The role of the sampling distribution in developing understanding of statistical inference.* Unpublished PhD thesis, Swinburne University of Technology, Melbourne.

Lipson, K. (2002). The role of computer based technology in developing understanding of the sampling distribution. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics*. [CD-ROM] Voorburg, The Netherlands: International Statistics Institute.

Marshall, S. P. (1995). *Schemas in problem solving*. Cambridge: Cambridge University Press.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1). Retrieved 28 January 2004 from the World Wide Web: http://www.amstat.org/publications/jse/v10n1/mills.html

Moore, D. S., & McCabe, G. P. (1999). *Introduction to the practice of statistics* (3rd ed.). New York: W. H. Freeman and Company.

Nitko, A. J., & Lane, S. (1990). Solving problems is not enough: Assessing and diagnosing the ways in which students organise statistical concepts. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics* (Vol. 1, pp. 467-474). Voorburg, The Netherlands: International Statistics Institute.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*, 937-949.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.

Piaget, J. (1970). *The principles of genetic epistemology*. New York: Columbia University Press.

Putnam, R. T., Lampert, M., & Peterson, P. L. (1990). Alternative perspectives on knowing mathematics in elementary schools. In C. B. Cazden (Ed.), *Review of research in education* (pp. 57-150). Washington, DC: American Educational Research Association.

Rubin, A. (1990). *Sampling laboratory* [Computer software]. Unpublished software.

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics* (Vol. 2, pp. 314-319). Voorburg, The Netherlands: International Statistics Institute.

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relations. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 91-104). Amsterdam: IOS Press and the International Statistics Institute.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Shavelson, R. J. (1993). *On concept maps as potential "authentic" assessments in science. Indirect approaches to knowledge representation of high school science*. Los Angeles, CA: National Centre for Research on Evaluation, Standards, and Student Testing.

Starr, M. L., & Krajcik, J. S. (1990). Concept maps as a heuristic for science curriculum development: Toward improvement in process and product. *Journal of Research in Science Teaching, 27*, 987-1000.

Stirling, D. (1991). *Models'n'Data* (Version 1) [Computer software]. Santa Barbara: Intellimation.

Stirling, D. (2002). *Computer Assisted Statistics Teaching* (Version 2.1) [Computer software]. For information about this software see http://cast.massey.ac.nz/.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110.

Williams, A. M. (1998). Students' understanding of the significance level concept. In L. Pereira-Mendoza (Ed.), *Proceedings of the 5th International Conference on Teaching Statistics* (Vol. 2, pp. 743-749). Voorburg, The Netherlands: International Statistics Institute.

Author

Kay Lipson, Swinburne University of Technology, Locked Bag 218, Lilydale, VIC 3140. Email: <klipson@swin.edu.au>

Appendix 1: Terms Supplied to Students for Constructing Concept Maps

Map		Terms
1	Sampling distribution for the sample proportion	centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable
2	Sampling distribution for the sample mean	centre, computer generated, constant, distribution, normal model, μ, $\bar{x}$, population, population mean, repetitions, sample mean, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable
3	Sampling distribution	centre, constant, μ, p, $\hat{p}$, population, population parameter, r, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable, $\bar{x}$
4	Hypothesis testing	alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic
5	Estimation	confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution
6	Statistical inference	confidence interval, decision and conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic