® THE IMPLICATIONS OF INTRODUCING REPORT WRITING INTO AN INTRODUCTORY STATISTICS SUBJECT

LIPSON, Kay and KOKONIS, Sue
Swinburne University of Technology
Australia

This paper reports the findings of a study undertaken by the authors to investigate the ability of students in an introductory statistics subject to undertake a report writing task. It was found that students find such a task quite difficult, and generally are able to score much higher marks on questions where they carry out analyses from first principles using a standard procedure. However, it is also suggested that the report writing task may be classified as a metacognitive activity, and of itself provides a means of facilitating the development of conceptual understanding in students.

INTRODUCTION

Statistics and Research Methods is an introductory statistics subject at Swinburne University of Technology, Lilydale. The subject has no mathematical pre-requisites, and students use the TI-83 graphics calculator for analyses. A goal of this subject has always been to develop both procedural understanding, the student's ability to carry out routine tasks successfully, and conceptual understanding, which implies an understanding by the student of what they are doing and why they are doing it (Hiebert & Carpenter, 1992). Assessments in the subject have been designed to reflect this goal.

In 2003 the decision was made to alter the emphasis of the statistical testing component of the subject. Prior to 2003 students were required to carry out hypothesis testing from raw data by following an essentially routine, procedural format using the calculator and to write up their findings according to a prescribed format of steps. An example of the standard question from the 2002 examination is given in Table 1.

Table 1
Example of Standard Question

A new super highway was built for mean speeds of 120 kph. It is suspected that over time the mean speed has increased. A sample of 12 cars was timed over 10 kilometres and their speeds (in kph) are given below. Do these data provide sufficient evidence to conclude that the mean speed now exceeds the original planned limit of 120 kph?

128	140	123	145	102	138
160	120	139	134	148	156

In 2003 a new style of question was developed to give the students experience in developing the writing skills necessary in later statistics subjects, where analyses are carried out using SPSS, and the emphasis is on report writing. In the new style of question the students are not required to carry out the analyses. Rather, the data, together with the calculator output which correctly addressed the problem scenario, is supplied. Students then use this output to write a brief report, which is a brief, non-technical summary of the statistical analyses including the following features:

1. An introduction which describes the scenario, gives the sample size, and states the question.
2. A description of the sample results, supported by the appropriate sample statistics.
3. The type of test carried out and the appropriate statistics.
4. Whether or not the results were significant.
5. A conclusion which relates back to the question.

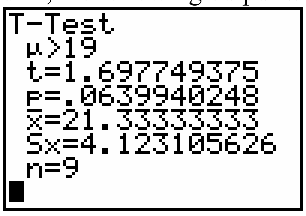
An example of the new style of question is given in Table 2.

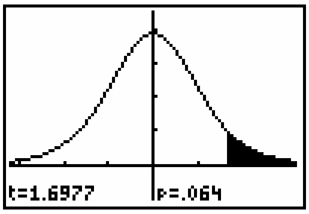
Table 2
Example of New Question

A mobile phone company wants to know if its marketing promotion for text messages has been successful. The average number of text messages sent per week per person prior to the campaign was known to be 19. A random sample of 9 customers is surveyed after the campaign, with each customer recording the following number of SMS messages:

21 27 23 20 18 19 24 26 14

Using the above data, the following output was produced:





Write a brief report summarising these results.

COMPARISON OF QUESTION STYLES

The standard style question requires students to decide on the basis of the scenario presented which of the seven hypothesis tests they have been taught is the appropriate test, to enter the data into their calculator, to use the calculator to carry out the test, and to present their answer according to a very structured answer format (hypotheses, level of significance, test statistic, p-value, decision and conclusion). Selection of which test to use can be challenging for students, but once the test has been selected the question can be considered a routine task.

The reaction of some of those concerned with the teaching and assessment of this subject was that the new style of question was far too easy, since the students were not required to either select the test or carry out the analyses for themselves. However, others felt that the new style of question had value as a metacognitive activity which may enhance students understanding.

Since the 1970's there has been a strong level of interest in the use of metacognitive activities to promote learning particularly from educators holding a constructivist view of learning. Metacognition can be defined as the knowledge and control one has over one's thinking and learning activities (Swanson, 1990). In an overview of how writing influences cognitive processes Klein (1999) found that analytical writing tasks, where students identified the goal of writing to support a claim and implemented elaborative strategies to do so, promoted conceptual understanding.

Metacognitive learning strategies used predominantly amongst language based arts subjects have been increasingly adopted by science educators (e.g. Yore & Hand, 2003) and mathematics educators (e.g. Kramarski, Mevarech, & Arami, 2002; Pugalee, 2004) with writing viewed as a tool that enhances metacognitive processes for problem solving. The use of writing activities in statistics has also been recommended as an activity which has the potential to develop understanding by many statistics educators (Dunn, 2000; Garfield, 2002; Stromberg & Ramanathan, 1996). This contrasts with earlier work by Beins (1993) who conducted a research project addressing the effect of introducing writing assignments in statistics. He found that including a writing task where students were required to summarise their statistical analyses in the form of a press release which did not contain technical language improved their ability in calculation and interpretation, but did not affect their level of conceptual understanding.

To investigate this further, the researchers were interested in the role of report writing as a metacognitive activity in this context and its implications for students' understanding. The report-writing task shown in Table 2 requires students to extract the appropriate information from the calculations given, interpret the sample statistics, provide a claim (that is, whether the results were significant or not) and give supporting evidence. This is consistent with the definition of analytical writing suggested by Klein (1999), suggesting that the report writing process could be classified as a metacognitive activity.

RESEARCH QUESTIONS

Previous studies have showed that the levels of procedural and conceptual understanding of statistical inference demonstrated by students can be quite different (Lipson, 1995). Since an important change had been made in the emphasis of the subject and its assessment, the researchers carried out a study to address the following research questions:

- How well were students able to undertake the statistical reporting task?
- Did the change in emphasis of the subject to include report writing have any implications for the students' procedural understanding?
- Did the change in emphasis of the subject to include report writing have any implications for the students' conceptual understanding?

METHODOLOGY

The data collected to address the research questions were based on the student responses to the examination questions in 2002 and 2003. Whilst these were clearly different students, they were similar cohorts in that they were both first year business students. Since the subject assessment is designed to measure both procedural and conceptual understanding, there were questions designed to measure both on each paper.

Students' performance on the standard question t –test from the 2002 paper (Table 1) was compared with a t-test question emphasising report writing from the 2003 paper (Table 2). Whilst superficially assessing the same topic, the students' knowledge of the t-test, comparison of the students' results in these questions would allow their relative performance in each task to be compared. A qualitative study of student responses to the report-writing question in Table 2 was also undertaken. Together these analyses address the first research question, concerning how well the students were able to carry out the report-writing task.

In order to address the second research question the scores achieved by students on a question which appeared in similar forms on both the 2002 and 2003 examination papers was compared. This question was concerned with the students' ability to carry out an hypothesis test for a single proportion.

As previously stated, the subject is designed to develop both conceptual and procedural understanding in students, and assessments have been designed to reflect this goal. Both papers had several items designed to address conceptual understanding. An item designed to measure conceptual understanding was selected from each paper for comparison, and these are given in Table 3. The item selected from the 2003 paper follows the question given in Table 2.

Table 3
Questions Used For Analysis of Conceptual Understanding

Question from the 2002 examination paper The percentage of people in a certain community who smoke was known to be 30%. After an extensive anti-smoking campaign was run in the community, a random sample of 100 people was surveyed and of these only 20% was smokers. The p-value for this result is 0.015. Explain what the p-value is measuring by completing the following sentence: The p-value is the probability that…
Question from the 2003 examination paper How would you explain (in non technical terms) to the manager of the store that the difference between the sample mean $\bar{x} = 21$ and the population mean $\mu = 19$ does not provide evidence that the campaign has been successful?

The student responses to these questions were analysed and coded according to a classification which arose from the analysis of their responses, details of which are given in the next section.

RESULTS

The analyses in this section are based on the final examinations undertaken by 226 students in 2002, and 266 students in 2003.

1. RESEARCH QUESTION 1

1.1 Analysis of student responses to the report-writing task

The report-writing task required students to describe the sample, indicate which test was used, decide whether a significant result had been found and provide supporting statistics (in this case the t statistic and p-value). The task was scored on a scale from 0 to 5, based on the requirements of a report outlined in the introduction. A histogram of the scores is given in Figure 1, showing that the distribution was approximately symmetric, with a mean score of 2.2, and a standard deviation of 1.3.

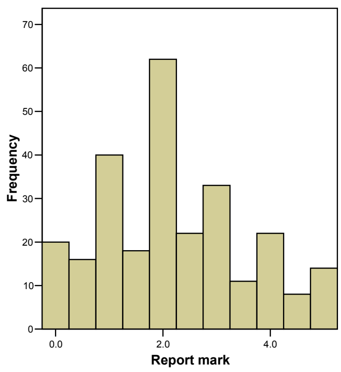


Figure 1. *Histogram of the student scores on the report-writing task.*

These scores indicate that, for many students the report writing task was quite difficult, with very few students awarded a mark of 4 or more. A qualitative analysis of the students' answers revealed that in writing the report the majority of students relied to some extent on the same terminology they had used for more procedural questions, with answers such as:

> *As the p value of .064 is greater than .05, we can say there is insufficient evidence to suggest the number of SMS messages has increased.*

This clearly does not fit the criteria of a report.

A few students endeavoured to explain their results in more conceptual terms such as sampling variability or probability of the sample result occurring, rather than to write a report. For example:

> *From the results above we can see that from a sample of 9 people the sample mean* $\overline{X}$ *= 21.33. The sample standard deviation is then* s_x *= 4.12. Giving a p value of p = 0.064. From the graph and the t value, we can see how many standard deviations it (is) away from the stated population mean of 19.*

Only a very small proportion of students (just over 15%) gave answers that addressed the criteria of report writing as they had been instructed. The following example would have scored full marks on this task:

> *A random sample of 9 mobile phone customers sent an average of 21 messages with a standard deviation of 4.12. While this was more than the average of 19 per week sent before a marketing campaign, a t-test found there was insufficient evidence to suggest the marketing campaign had been successful (t(8) = 1.698, p = 0.064).*

In summary, even though students had been given many examples of reports in their notes and sample assessment tasks, the report writing task proved to be quite challenging for this cohort of students, and not quite the straightforward question that some had anticipated.

1.2 Comparison of scores on the t-test questions in 2002 and 2003

For comparison the scores for the standard t-test question from 2002 and the report writing question from 2003 were scaled so that the scores represent the proportion of available marks. Boxplots showing these scores are given in Figure 2.

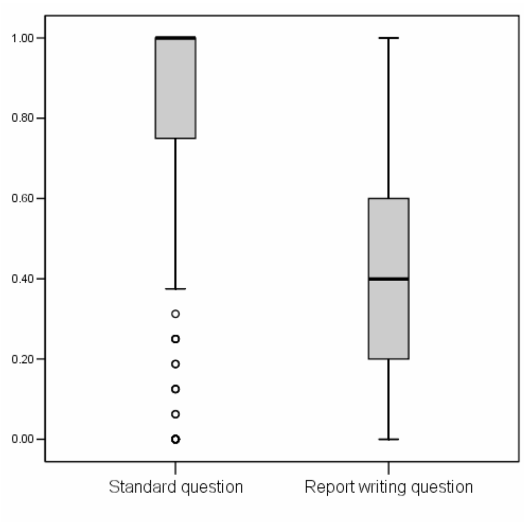


Figure 2. *Boxplots of the student scores on the standard question 2002 and the report writing question 2003.*

As can be seen from the boxplots, the students' scores were much higher on the standard t-test question, with a median score of 100% of the available marks, compared to a median score of only 40% of the available marks on the report writing question. In fact, apart from the six students who scored poorly in 2002, all of the 2002 scores were higher than the 2003 median. This comparison supports the previous theoretical analysis that rather than making the question much easier by giving the students the test and the analysis results, the report writing style of question proved to be demanding for the students.

2. RESEARCH QUESTION 2

A question which had appeared often in previous examination papers and a form of which was given on both the 2002 and 2003 examination papers was used as a measure of procedural understanding for each group of students. The maximum mark obtainable for the question was 13. Boxplots of the scores for each cohort of students, 2002 and 2003, are given in Figure 3.

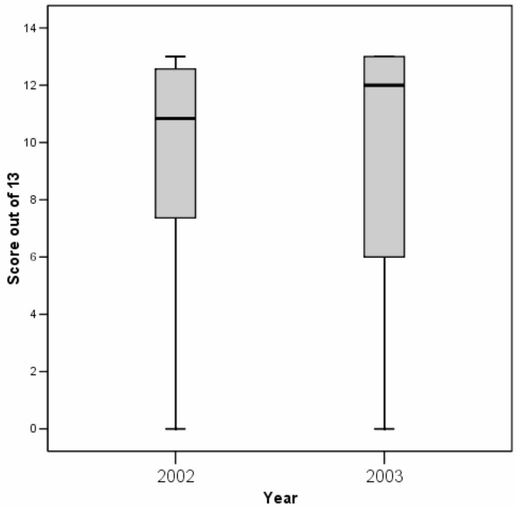


Figure 3. *Boxplots of the student scores on the same standard question 2002 and 2003.*

From Figure 3 it can be seen that on the whole the students scored very well on this question. The distribution of marks for both groups is extremely negatively skewed, with the median mark in 2002 of 11 and the median mark in 2003 of 12. Whilst there is an increase in median mark, it is relatively small. Thus on the basis of this data it would appear that the change in emphasis of the syllabus has had little effect on the students' procedural understanding. This is perhaps what could reasonably be expected, since the students were already performing well on this task.

3. RESEARCH QUESTION 3

A qualitative analysis of student responses to the questions designed to measure conceptual understanding (see Table 3) was carried out. To do this, three researchers independently reviewed the students' answers and categorised them according to common themes. Up to seven separate themes were identified by individual researchers, but after discussion and comparison the responses were further classified into three broader themes:

- Non-attempt, unintelligible or wrong.
- Responses where the student indicated either correctly or incorrectly the actions which they would undertake when presented with the p-value, rather than an attempt to interpret the p-value. These answers were classified as *operationalised*.
- Responses which correctly referred to sampling variability, sample size, the chance of the sample occurring, the relationship between the sample and population, the size of the difference between the sample statistic and the population parameter, sample size, or a reasonable combination of some of these. These answers were classified as *at least partially correct*.

This classification allowed the student responses from 2002 and 2003 to be analysed and compared. The percentages of answers in each category in each year are given in Table 4.

Table 4
Classification of Conceptual Understanding Questions

	2002 (%)	*2003 (%)*
Wrong	25.1	27.4
Operationalised	27.6	6.0
At least partially correct	47.2	66.5

3.1 Analysis of Conceptual Understanding Question 2002

Approximately one quarter (25.1%) of student responses were identified as a non attempt, unintelligible or as clearly wrong with answers such as:

> *…1.5 out the 100 people surveyed smoke.*

Another quarter (27.6%) of students operationalised the answer instead of explaining what the p value was measuring. This was done successfully by 23.8% of students, with answers such as:

> *…the anti-smoking campaign has worked because the p-value is less than 0.05, therefore we reject Ho (or 30% who smoke) and conclude that the community now has 20% smokers (reject Ho and accept that H1 is less than the null hypothesis). eg. 20%.*

and unsuccessfully by 3.9% of students, for example:

> *…the percentage of people in a certain community who smoke has not decreased since the anti-smoking campaign was run in the community.*

Just under half (47.2%) of students gave responses that were at least partially correct. Examples of responses in this category included:

> *…we could find a random sample of 100 people producing a result of 20% being smokers. A 0.015% chance (very small) is the chance of finding a sample as extreme and similar to this.*

Of the responses in the last category 20% were considered to be a good interpretation of the p-value, such as:

> *…the p-value is the probability of attaining 20% smokers in a sample of 100 that came from a population where 30% of people were smokers.*

These answers indicate that the majority of students were not able to articulate the meaning of the p-value in non-technical terms.

3.2 Analysis of conceptual understanding question 2003

Of the student responses to the question designed to measure conceptual understanding on the 2003 examination paper, 27.4% were classified as non-attempt, unintelligible or wrong. Those which were clearly wrong included answers such as:

> *…The difference between the sample mean and the population mean does not provide evidence that the campaign has been successful because the population standard deviation is unknown therefore, the sample proportion has been used to conclude the results.*

Only a very small group of students (6.0%) gave an answer which was classified as operational. An example of an answer classified under this category is:

> *…You would explain that you would like to have a significance level of 0.05 and assess it with a p-value, using your hypothesis testing (z test).*

By far the largest group of responses (66.5%) was classified as at least partially correct. To be classified in this group, students generally mentioned sampling variability, sample size, size of the difference between the sample mean and the population mean, or any combination of these three. Examples of answers included in this category are:

> *…Because the mean difference between the sample and the population is not great, it can be attributed to sampling variability. The difference is not great enough to be able to provide evidence that the campaign has been successful.*

and

> *…Only a very small sample of 9 was given this is not enough sample data to make a clear forecast of the campaign if more sample data was given a better indication could be given.*

and

> *…Because the difference between the means is so low we can't be sure it was caused by the ad campaign and not variations in the sample.*

These answers indicate that many of the students were able to communicate a level of conceptual understanding of the formal hypothesis testing process.

DISCUSSION

Overall the students found the report writing task more challenging than the structured style of question presented the previous year. While the majority of students still relied on terminology they had employed for more procedural questions, most of them made an attempt to at least reorganise the information. Whether this indicated a genuine attempt by students to interpret the information at a deeper level of understanding or was simply a reshuffling of information learnt at a procedural level is unclear. Klein (1999) suggests three critical junctures need to be satisfied for analytical writing to be regarded as fulfilling a metacognitive function. Firstly writers must adopt the goal of writing as an analytical task. Secondly they must adopt elaborative reasoning operations to achieve this goal and thirdly these operations need to be of a sufficient level to bring about substantive and valid changes to their knowledge. The nature of this study did not provide us with the scope to ascertain whether these critical phases were occurring, however preliminary results indicate that this could indeed be the case. On the basis of the analysis of the responses to the understanding questions, there was a higher level of conceptual understanding demonstrated by the students in 2003 than in 2002. This suggests that the shift in emphasis to report writing may have facilitated understanding at a metacognitive level, however further research into the cognitive processes taking place during report writing in this context needs to be carried out before any valid conclusions can be reached.

The emphasis on report writing in 2003 had very little impact on student's procedural understanding with no significant difference recorded between similar questions on the 2002 and 2003 papers. As there had been very little change to the teaching of this style of question and previous students had already demonstrated a level of proficiency with this type of procedural question, this finding was not unexpected.

LIMITATIONS

The above study was not intended to be definitive but exploratory and the results need to be viewed in the light of the following limitations. Firstly the design was observational not experimental so other important variables such as changes in teaching staff and differences in the student cohort between the two years were not controlled for. In addition the questions used to assess conceptual understanding differed in emphasis across the two years. The question in the 2002 paper more readily lent itself to interpretation from an operational perspective which may account for the large proportion of students who answered it in this manner.

IMPLICATIONS FOR TEACHING AND ASSESSING

Although the primary motivation for the changes in focus and assessment documented here was to introduce report writing to students as a precursor for the more extensive demands of report writing in later years, the benefits of report writing as a metacognitive activity may have implications for the further development of the subject. The challenge now is to engage students in report writing as an analytical writing process that provides opportunities for elaborative reasoning skills to be developed. By providing students with learning opportunities and examples to achieve theses goals there is also the danger that report writing may become proceduralised if used often enough. However, there is preliminary evidence here that the challenge of the report writing task has encouraged students to take an holistic view of the processes that underlie statistical inference, and report writing can thus be classified as a metacognitive activity which facilitates the development of conceptual understanding.

REFERENCES

Beins, B.C. (1993). Writing assignments in statistics classes encourage students to learn interpretation. *Teaching of Psychology, 20*(3), 161-164.

Dunn, D.S. (2000). Letter exchanges on statistics and research methods: Writing, responding and learning. *Teaching of Psychology, 27*(2), 128-130.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3), http:www.amstat.org/publications/jse/v10n13/garfield.html.
Hiebert, J., & Carpenter, T.P. (1992). Learning and Teaching with Understanding. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-97). New York: MacMillan.
Klein, P.D. (1999). Reopening inquiry into cognitive processes in Writing-to-Learn. *Educational Psychology Review, 11*(3), 203-272.
Kramarski, B., Mevarech, Z., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*, 225-250.
Lipson, K. (1995). Assessing Understanding in Statistics. In J. Garfield (Ed.), *Collected Research Papers from the 4th International Conference on Teaching Statistics*: International Statistics Institute.
Pugalee, D. K. (2004). A comparison of verbal and writtend descriptions of students' problem solving processes. *Educational Studies in Mathematics, 55*, 27-47.
Stromberg, A. J., & Ramanathan, S. (1996). Easy implementation of writing in introductory statistics courses. *The American Statistician, 50*(2), 159-164.
Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82*(2), 306-314.
Yore, L. D., & Hand, B. D. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*(6), 689-725.