Chapter Ten
A Re-Analysis of the Second Exploratory Study

The student responses to statistically based information in the second exploratory study are re-examined using judgement criteria developed from our statistical thinking framework. The questions and student responses are discussed, including their implications for teaching. The students' current and previous exposure to the statistical discipline have not engendered or embedded a depth and quality to their statistical thinking. It is conjectured that students need some scaffolding, such as the judgement criteria, in order to evaluate statistically based information, and need practise at reasoning with data and context. Our statistical thinking framework is considered to be useful for developing the judgement criteria.

10.1 Introduction

When students are asked to interpret and evaluate statistically based information it is not clear to statistics educators the type of answers that should be expected (Friel et al., 1997). This is borne out by Gal et al. (1995) whose paper resulted from a two hour working group discussion on the assessment of interpretive skills. Much of the discussion centred on the definition of interpretive skills, and thus is based on conjectures of what skills they thought students might need for interpreting media reports. They proposed the development of a "critical list" of questions that students should invoke when evaluating information. The assessment of interpretive skills was briefly addressed. However it was noted that judgements of student responses would be subjective, and that scoring rubrics would be needed but *"generic scoring rubrics cannot easily provide information about what specific questions on the "critical list" are not being invoked by the student, and thus what needs to be emphasised in instruction"* (p. 25).

Our proposed statistical framework was created from an ongoing analysis and interpretation of data from four exploratory studies. It would be now useful to return to the second exploratory study and check whether the research findings could be used to articulate the type of answer that would be desirable from a student when presented with statistically based information. Thus the framework will be used to explicate the judgement criteria or "critical list" necessary for evaluation. The criteria will then be employed to generate solutions to some of the items presented in the second exploratory study. These normative solutions will be put into a scoring rubric and the statistical thinking of the students will be re-examined. This should lead to better detection of the type of thinking that is being operationalised in the students when they are presented with the items. This examination of the second exploratory study items should also provide some clearer guidelines for instruction.

10.2 The Interpretation of Statistically Based Information

Before determining the types of questions that should be asked in the evaluation of statistical information, the purpose of such information and the role of the student reader/listener should be addressed in order to define the domain of the questions.

Purpose of a Statistically Based Report

There are many types of statistically based reports, from formal research reports to reports for non-statistical clients, to media reports. The level of sophistication will depend on the intended audience. A statistically based media report is usually presented because someone has noticed a pattern in data, or wishes to "popularise" research results for the general public. Hence the writer of the report is "selling" or purporting to the readers/listeners that they should take cognisance of the conclusion or inferred conclusion.

Role of the Student Reader/Listener

When student readers/listeners are presented with a statistically based report we want them to be able to give a critical appraisal. They need to determine what the report is wanting them to believe, or trying to suggest to them. After considerations based on their statistical and contextual knowledge, they must weigh up what they are willing to believe, what else should be done, or what should be presented to them to convince them further. We, and perhaps other statistics educators, would like to enable the students to make valid judgements and be able to articulate the reasons for their judgement.

Relationship to the Statistical Thinking Framework Dimensions

When reading and attempting to understand a statistically based report, a summary of the key points is a helpful start for clarification of the main issues that are being addressed in the article. But a summary of the key points is insufficient if we want students to critically appraise the report. Identification of the stages of the investigative cycle dimension (PPDAC) would facilitate a thorough and systematic evaluation. Once the stages of the cycle have been identified, then each can be critiqued with pertinent questions addressed at definition issues, measurement issues, design issues and communication issues (see Section 9.4 for trigger questions on measurement). During the interpretation of statistically based information, the readers/listeners move into the interrogative cycle dimension. Interrogative cycles are in active use (see Section 8.4.1) during evaluation. For every issue being appraised there is first ‘seeking information’ from the article, ‘generating possibilities’ raised by the issue, ‘interpreting’ and ‘criticising and evaluating’ the issue in the article. The last step is the ‘judgement’ of the issue in the article. While each stage of the investigative cycle dimension is being critiqued, the types-of-thinking dimension is also operationalised. These appear in the list of questions below (Section 10.3) albeit under different nomenclature. The statistical thinking framework is a generalised structure, and therefore the general principles were used to address the specialised task of interpretation. For example ‘transnumeration’ appears in the ‘measurement’, ‘rethinking the analysis’ and ‘communication of data’ categories. ‘My own internal check’ is an example of ‘reasoning with statistical models’ and recalling ‘techniques’ while ‘seeking explanations’ is under ‘variation from a context perspective’, ‘causal validity’ and ‘external checks’. The dispositions dimension which affects entry into a thinking mode has been discussed also in Section 8.4.1.

10.3 The Judgement Criteria

We present judgement criteria or a “critical list” in a form that could be used by students to aid the articulation of the reasons for their judgement. Their reasoning could take into account, for example, whether the postulated or inferred claim is more plausible than other possibilities, or whether the report has demonstrated convincingly that there is a pattern in the data. In order to build up *investigative skills* in statistics students, we believe that it is important for them to go beyond the stage of judgement. They should be able to generate ideas for what they would do next if they were the investigator. This is reflected in the criteria.

The decision as to whether the conclusion is valid means addressing the following main questions or issues:

Conclusion Validity

- *Do the data and arguments address the problem?*
- *Is the claimed inference space valid?*
- *Was sampling variation properly taken into account?*
- *Were other important sources of variation properly taken into account?*
- *Are the arguments and recommendations logically consistent?*

Underlying these main issues the following specific questions, based on the investigative cycle and the types of thinking dimensions, should be considered to form a judgement about the presented information.

Problem Validity

- *What problems/questions are being addressed?*
- *Why are these questions being addressed?*
- *Are these questions valid and useful in the system under study?*

Measurement Validity

- *Have the "right things" been measured in the "right way"? (i.e. Do the measures really address the problem?)* Measuring the wrong things is fatal to any argument!

Data Capture Validity

- *Have the data actually been obtained from the population/process the conclusions have been applied to, or from something else (e.g. non-representative sub-population or different population)?*
- *Has a reasonable "sampling mechanism" been used?*
- *Is the non-response rate (or data-missingness rate) too high for the results to be credible?*
- *How much variation is introduced by the measurement process?*
- *Have reliable data collection and management procedures been used?*

These last three questions will depend on the type of report being evaluated. In ‘popular’ articles you may need to trust the author. Nevertheless you should still consider such questions and use any context knowledge you may have about the situation.

Analysis Validity

Basic

- *Is the analysis appropriate for the types of measurements used?*
- *Is the analysis appropriate for the method of data-collection used?*
- *Is there any assurance that models used in the analysis actually fitted the data?*
- *Are the statistical models (graphs, statistical summaries etc.) appropriate for the data and problem?*

My Own Internal Check

- *What information can I comprehend/extract from the statistical models (graphs, statistical summaries etc)?*
- *Are there things that I know which allow me to do a quick plausibility check on the information? (e.g. Is there a statistical model or other model that I know about which I can use to check this information?)*
- *Are there things that I know from a context perspective that contradict this information?*

Variation

from a statistical perspective

- *Was sampling variation properly taken into account?*
- *Is this special cause or common cause variation?*
- *Are margins of error/confidence intervals given to indicate levels of uncertainty in all the estimates quoted? (If not, can I construct them myself from the information?)*

from a context perspective

- *Were other important sources of variation (known or potential) relating to this information taken into account?*

Rethinking the analysis

- *Is there another way of looking at the data which may reveal more about the situation?*

Causal Validity

(for evaluating causal conclusions drawn)

- *Have all competing explanations that seem plausible been taken into account/investigated?*
- *Will the "cause" be useful in practice (is it something that can be altered to improve a situation), or is it at best a pointer to a deeper cause?*
- *Have the judgement criteria (see epidemiology criteria in Section 9.3.3) been taken into account?*
- *Is the proffered cause special to this population/environment or is it possible it could apply more generally?*

Communication of Data

- *Is there anything about the way the data are presented which is misleading? (Note: a misleading graph does not imply that the claims or conclusions in the report are incorrect, or that the reader cannot compensate and reach correct ones.)*
- *Have the data and findings been communicated in a way that is readily understandable?*

External Checks

- *Has the study and its conclusions been scrutinised by experts in the field? (e.g. published in a reputable research journal and have undergone peer review)*
- *Is the author a reputable researcher in the field? (Am I willing to accept the authority of the source of this article?)*

- *Are the sources for this report reputable? (e.g. Science versus Reader's Digest.)*
- *Are there any hints that the ideas may have been mistranslated from reputable sources? (e.g. journalists writing from research articles that they have not properly understood.)*

From these judgement criteria I generated possible solutions to six of the second exploratory study items. These solutions are formed into a simple three point scoring rubric (see Appendix Five for solutions to the six questions). I subsequently assessed the students' responses in conjunction with my supervisor. The six students are referred to as JOY, NORS, MORTA, EAGLE, ISA and TEP (their own pseudonyms). No definitive conclusions can be formed from the resultant data, since the students' responses depended upon interview technique and subjective marking or interpretation of their responses. Nevertheless some interesting patterns emerged which may cause some reflection on the teaching of statistics.

The following should be noted about the scoring method:

1. In the criterion 'variation from a context perspective' a score of 1 is given if the student produced only individual-based explanations or factors pertaining to an individual or group of people. Whereas a score of 2 is given for also mentioning system explanations or factors pertaining to the system under question.
2. Not all questions operationalised all the criteria and therefore these criteria are left as a blank in the presented data. The students TEP and NORS have no data for the 'fitness newspaper question' since TEP's tape tangled during transcription and time ran out in NORS's interview.
3. Apart from the conclusion criterion, a blank square indicates unable to be ascertained from the interview, a score of 0 means not observed or absent, a score of 1 indicates the student has some rudimentary ideas in the particular criterion, whilst a score of 2 indicates some ideas are present though they may not be well articulated. For the conclusion criterion, 0 means incorrect conclusion, 1 indicates a partially correct conclusion whereas 2 indicates a somewhat correct conclusion, though perhaps not fully articulated and ideas not fully formulated.
4. The criteria questions tend to overlap and could possibly be collapsed into fewer criteria but for the purposes of this analysis all criteria seemed to be needed to form a full judgement on the data presented.
5. The use of these criteria questions to produce a scoring rubric forced me to think in specific ways, and thus facilitated a much deeper analysis than I had been capable of previously. For example in the 'prison newspaper question 4' I had not considered statistical variation in relation to the ratio 1.5 before, nor was I aware of the depth of thinking that was able to be generated from such a small phrase (see Appendix Five).

The students' responses were also used in the development of more aspects to the solutions. There are aspects that will not be covered in the solutions but this would not affect the scoring rubric used.

6. The students did not have the judgement criteria nor had they been exposed to any method for framing their responses for evaluation.

10.4 Analysis of Responses to Questions

From an analysis of the responses to six questions (see Figs. 10.1 - 10.6 below and tables of data in Appendix Five), some patterns emerged in regard to the 'investigative cycle', 'statistical variation', 'transnumeration', 'reasoning with models', 'seeking explanations' and 'conclusions'.

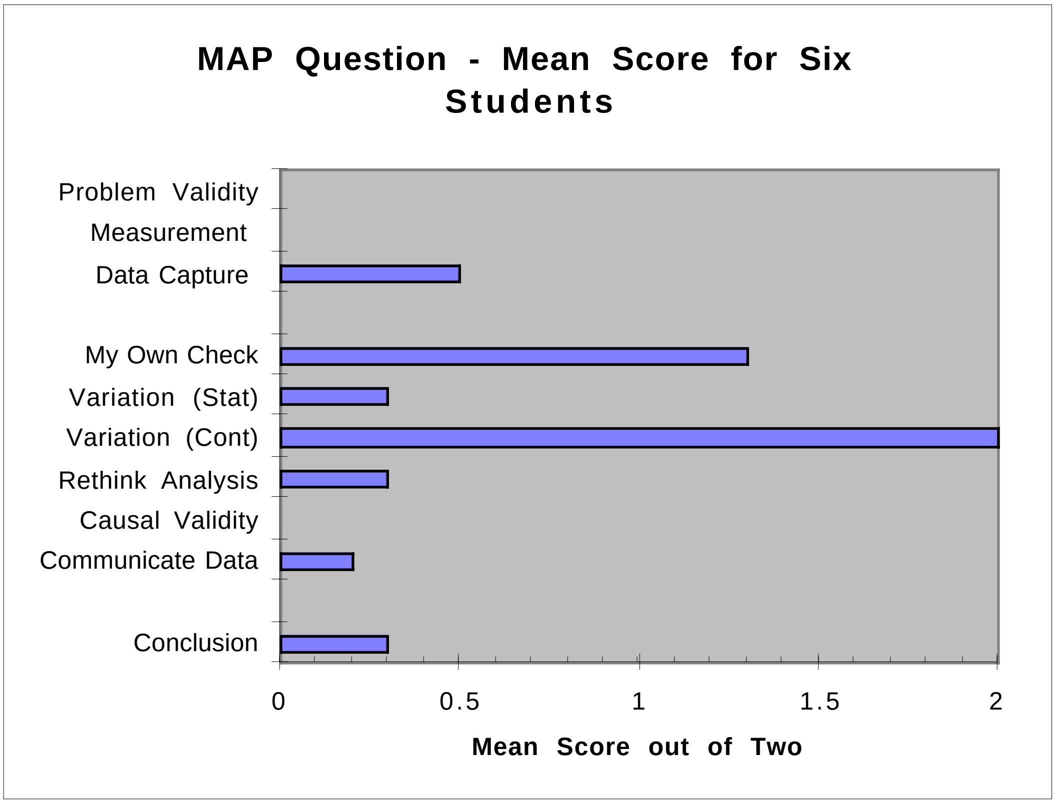

Figure 10.1 Student Responses to Map Question
[Note: Difference in responses to statistical variation and context variation]

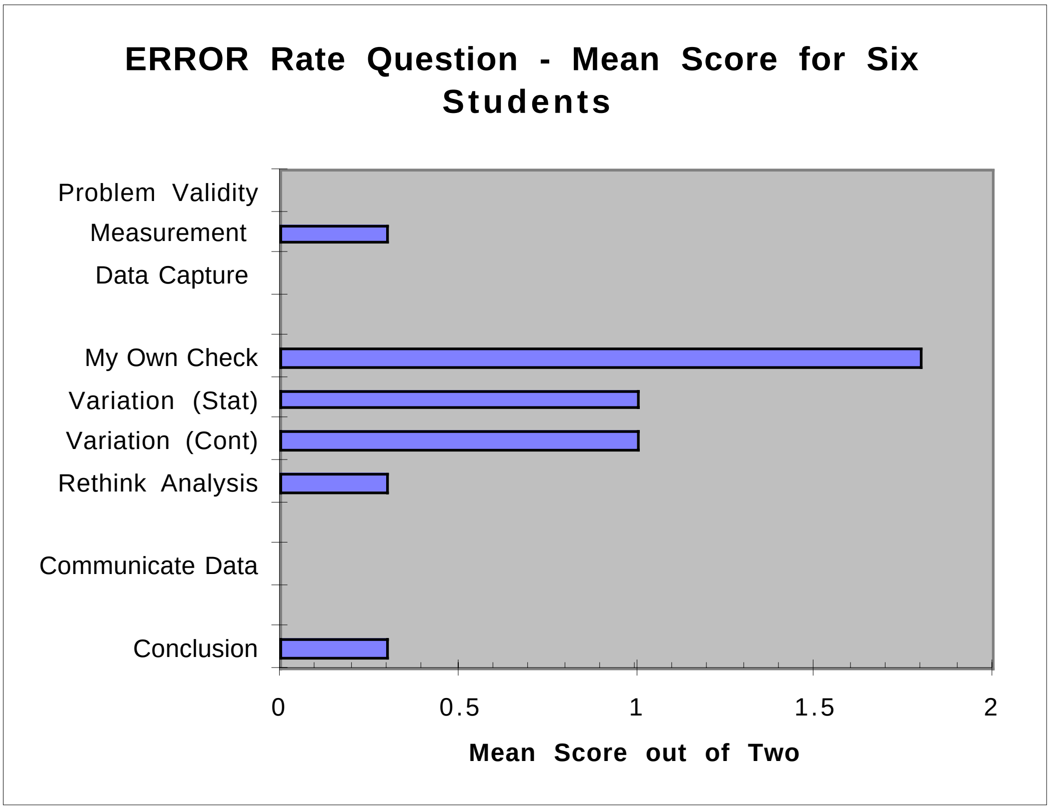

Figure 10.2 Student Responses to Error Rate Question
[Note: Same response to statistical variation and context variation when data presented in boxplots]

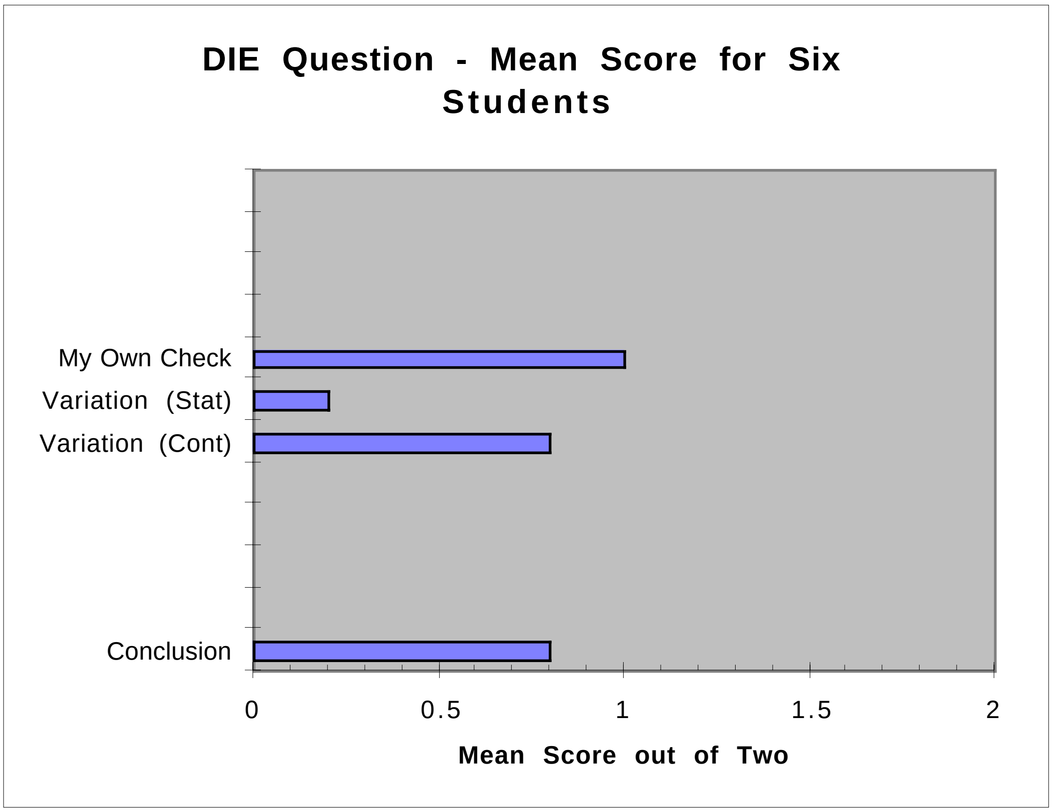

Figure 10.3 Student Responses to Die Question
[Note: Little response to statistical variation]

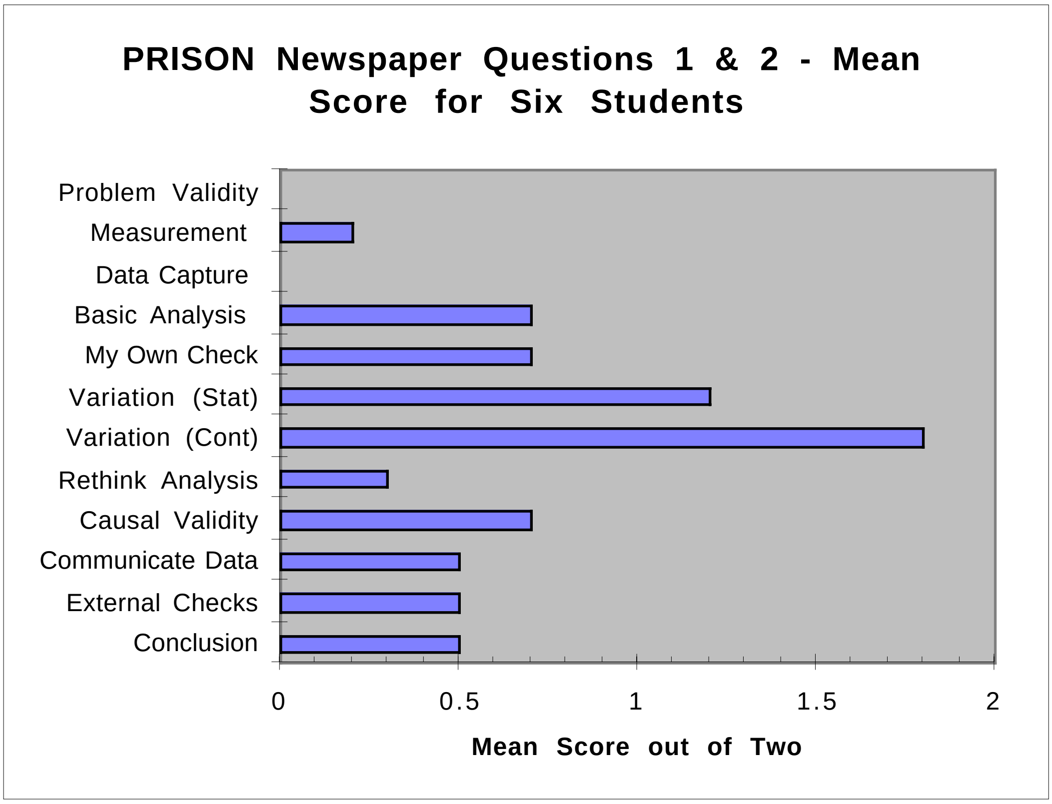

Figure 10.4 Student Responses to Prison Newspaper Questions 1 & 2
[Note: Similar response to statistical variation and context variation when data presented in different time intervals]

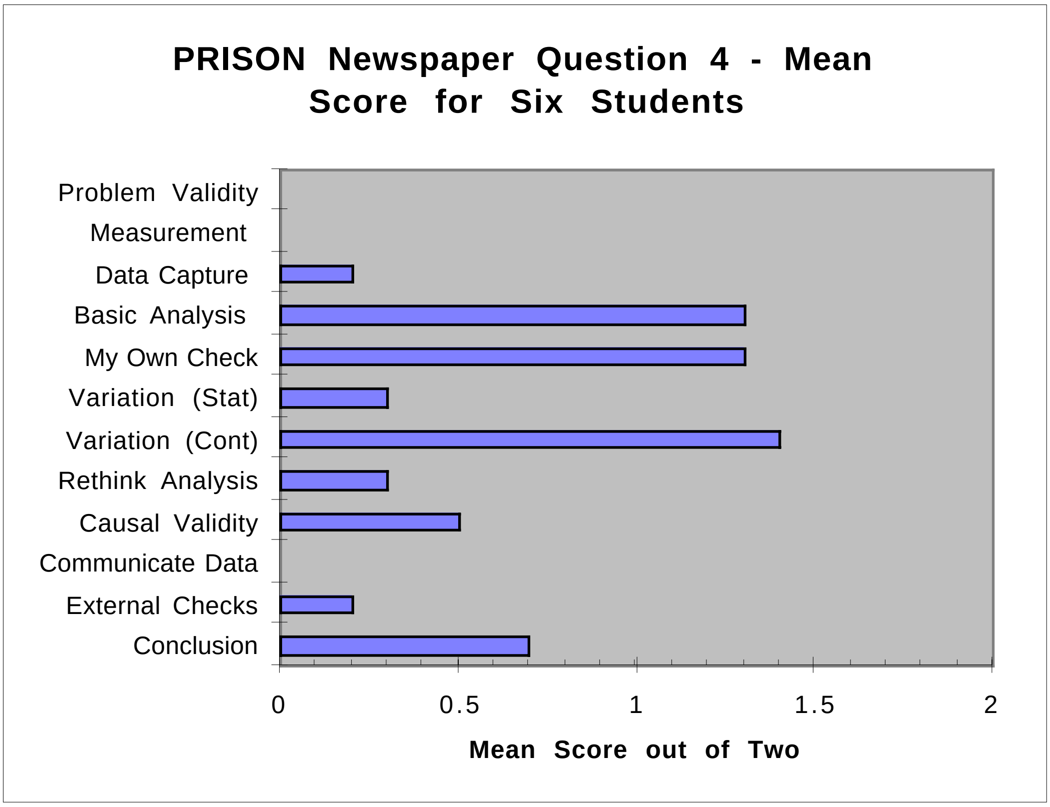

Figure 10.5 Student Responses to Prison Newspaper Question 4
[Note: Difference in responses to statistical variation and context variation]

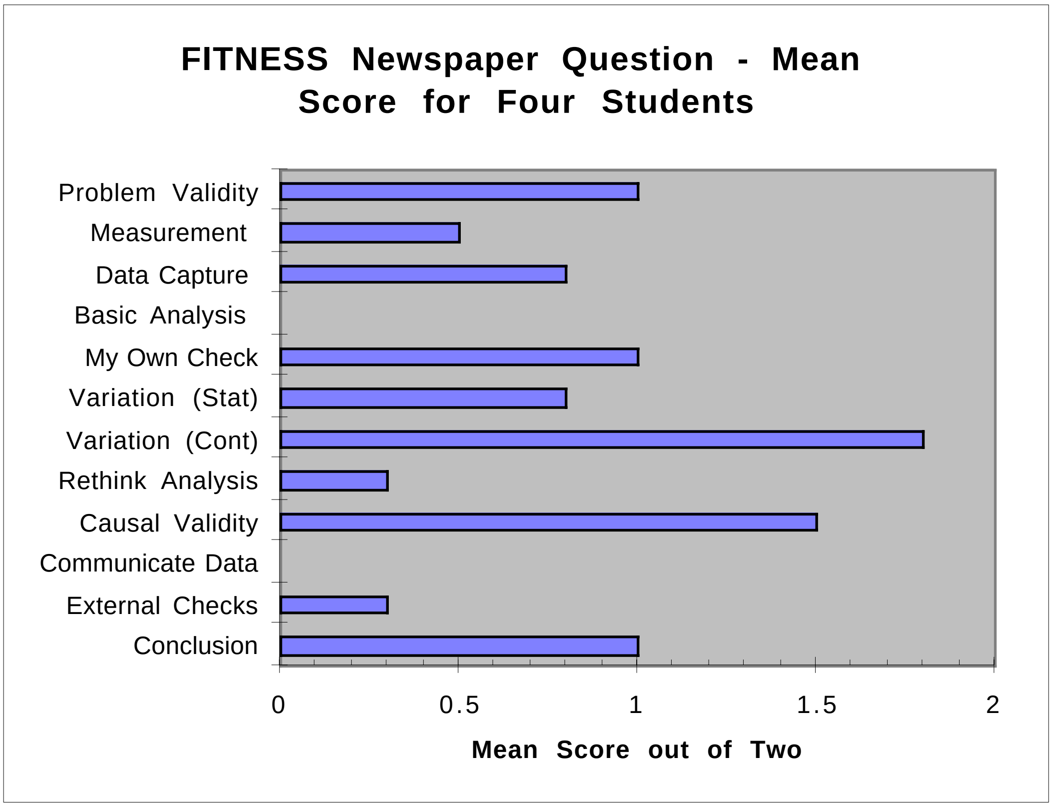

Figure 10.6 Student Responses to Fitness Newspaper Question

[Note: Extent to which each criterion is stimulated in article, and the good coverage of the investigative cycle]

The Investigative Cycle

The students' responses in the investigative cycle dimension appear to depend on the information given in the question. They *react* to the information given in the article rather than identify gaps in the information. They fail to construct, from their own knowledge of the empirical enquiry process, concerns that should be raised. These concerns, that can be determined through a bringing together of their own knowledge and the information in the article, are not easily produced. For example in the 'fitness newspaper question' (Fig. 10.6) the notion of 'data capture validity' is triggered since the article states that data are obtained from two treadmill tests. 'Problem validity' is only stimulated in this question since the article portrays fitness as being desirable for them personally, and this may have engaged the students and their beliefs. However when a graph is presented the students' attention appears to be on that rather than the text of the article. In the 'error rate question' (Fig. 10.2), the text states the data are gathered over one month yet no questions were raised about 'data capture validity'.

Statistical Variation

The 'statistical variation' type of thinking is triggered mainly in the 'error rate question' (Fig. 10.2) and the 'prison newspaper questions 1 & 2' (Fig. 10.4). My conjecture is that because the 'error rate question' is presented as boxplots it would be expected that students will notice the variation. The other question presents the data in text, with numbers and averages over different time intervals, and thus students notice the varying numbers. It also seems easier for the students to conjure up a visual picture of variation for a times series plot. This image is sequential points varying up and down with a trend line on a graph.

Transnumeration

'Transnumeration' is rarely stimulated in any of the questions. 'Communication of data' is a common feature of statistics courses as students are taught how to construct graphs and are presented with graphs and asked to analyse why they are misleading. However when the data are in a text article, it does not occur to most of the students to state that a graph would be helpful for communication. Nor do they respond to a non-standard graph, such as in the 'map question' (Fig. 10.1), that a more accurate presentation is required. 'Measurement validity' and 'rethinking the analysis', or suggesting other ways to look at the data, are not criteria upon which the students' attention is generally focussed.

Reasoning with Models

The criterion 'my own internal check' is triggered quite strongly in all questions as students interpreted the claims, or inferred claims, from the information. They could presumably readily recall, read and understand some familiar basic statistical techniques and terminology such as boxplots, ratio and tossing a die. In two questions they were required to recognise the situation and apply a technique to further understand the situation. In the 'map question' (Fig. 10.1) most of the students successfully used their general everyday context knowledge of the New Zealand population distribution to interpret the data. However in the 'prison newspaper questions 1 & 2' (Fig. 10.4) they were unable to use their statistical knowledge to recognise and apply a possible Poisson distribution to interpret the data (cf Scholz, 1987). The 'basic analysis validity' criterion is stimulated in the 'prison newspaper question 4' (Fig. 10.5) since a ratio is given for a named group of people, and thus there is a reaction to the given information. In two questions the basic 'analysis validity criterion' is not triggered because, we believe, students are required to produce a new perspective on the data. For example, the students did not suggest that a rate might be a better method of analysis rather than a raw number in the 'prison newspaper questions 1 & 2' (Fig. 10.4), nor did they respond from a non-critical or affirmative perspective that the given rates in the 'fitness newspaper question' (Fig. 10.6) were appropriate, and that the raw numbers in each fitness category could be assumed to be different.

Seeking Explanations

'Variation from a context perspective' is triggered for every question suggesting that it is natural for students to look for causes, even in the case of tossing a die. The way the question is presented and the information contained in the question appear to have no effect on this criterion. Individual-based and system-based explanations are prompted, although not consistently, in most questions, except the ‘error rate question’ (Fig. 10.2). This suggests that the students perhaps lacked context knowledge about work situations and that the two different types of explanations are not part of their thinking. ‘Causal validity’ and ‘external checks’ criteria are largely prompted by the context of the article. In the ‘prison newspaper questions 1 & 2 & 4’ (Figs. 10.4 and 10.5) and the ‘fitness newspaper question’ (Fig. 10.6), the author or research organisation is given, and therefore it is not surprising that the students reacted to such information. In the ‘fitness newspaper question’ the fitness cause is the prominent message in the article and therefore a student reaction is triggered. For the ‘prison newspaper questions 1 & 2’ there is a prompt to proffer possibilities, or community “wisdom” for the cause of the high level of suicides in prison, and thus there is a response from the students.

Correct Conclusion

Only in the ‘fitness newspaper question’ (Fig. 10.6) is the conclusion mean score 1 or partially correct. (All the other questions resulted in a conclusion mean score below 1.) This question has the greatest amount of information with regard to the investigative cycle and the students are familiar with the context. Thus, through reacting to the given information, the students could produce a sounder conclusion.

10.5 Analysis of Individual Student Responses

In the analysis of individual student responses (see Tables 10.1-10.6 and Figs. 10.7-10.12), a summary of their mean scores in each criterion is given, to determine an informal rank of a student’s thinking in relation to the other students. It should be noted that the rank is for the thinking that was revealed in the interview, and that the nature of this study precludes any formal statistical analysis. ‘Statistical variation’ and ‘transnumeration’ are also considered for the individual students, in an attempt to find patterns in the data that had not emerged in the analysis of the questions.

JOY

JOY is a female student, in the 17-19 year age group. Her highest school level mathematics studied was Form 7 ‘Mathematics with Calculus’. She was enrolled in the Stage 1 statistics paper at the time of her participation in the study (see Chapter 6 for details of course covered).

JOY has a mean score: 1 or greater for 3 criteria; between 0.5 and 1 for 5 criteria; between 0.1 and 0.5 for 3 criteria and no response for 1 criterion (see Table 10.1 and Fig 10.7). These mean scores give JOY an informal rank of 2 (out of six people).

The 'die question' stimulated some ideas about statistical variation for JOY. This did not occur in any of the other student responses. Her ideas were based around long run relative frequency random-device graphs. The 'prison newspaper questions 1 & 2' produced some ideas about variation based around the visualisation of a time series plot from the text.

JOY is the second best student for the 'transnumeration' type of thinking which arises in the 'measurement', 'rethinking the analysis' and 'communication of data' criteria. She always read everything in detail and endeavoured to gain comprehension of the information. She is the only student that seemed to be able to think of other ways that the data could be analysed from a population perspective.

Table 10.1 Question Data for JOY

JOY	Prob Valid	Measurement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	0		0	1	2	2	0	1		0
ERROR	0	0	0		2	1	1	0		0		0
PRS1|2	0	0	0	2	1	1	2	0	1	1	2	0
PRS4	0	0	0	2	1	0	1	0	0	0	0	1
FIT	0	2	1	0	2	0	2	0	1	0	0	1
DIE					0	1	2					2
MEAN	0	0.4	0.2	1.3	1	0.7	1.7	0.4	0.5	0.5	0.7	0.7

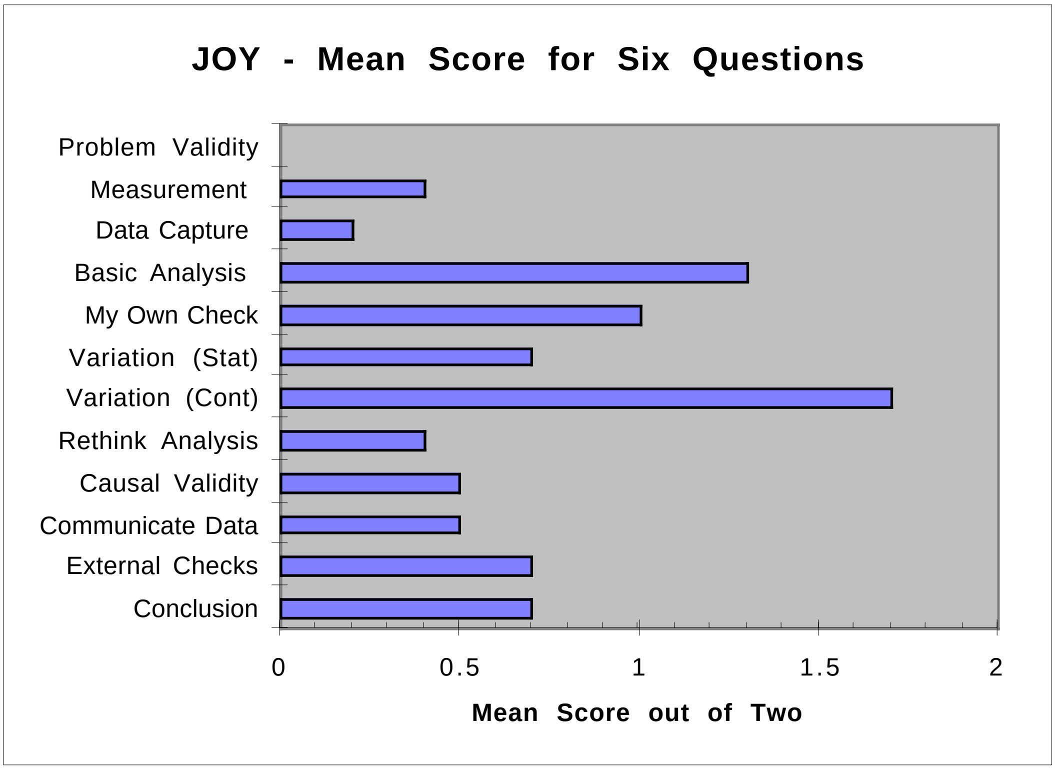

Figure 10.7 Criteria Responses for JOY
[Note the 3 criteria in which JOY scored 1 or more.]

NORS

NORS is a male student from an unskilled background, in the 40+ year age group. His highest school level mathematics studied was Form 5 Mathematics. He was enrolled in the Stage 1 statistics paper at the time of his participation in the study.

NORS has a mean score: 1 or greater for 6 criteria; between 0.5 and 1 for 4 criteria; between 0.1 and 0.5 for 1 criterion and no response for 1 criterion (see Table 10.2 and Fig 10.8). These mean scores give NORS an informal rank of 1(out of six people).

The ‘prison newspaper questions 1 & 2’ stimulated some well formed ideas about statistical variation, possibly because of the visualisation of a time series plot from the text. What is surprising is that NORS believed the die was biased in the ‘die question’ and gave no consideration to statistical variation.

NORS is the best student for the ‘transnumeration’ type of thinking which arises in the ‘measurement’, ‘rethinking the analysis’ and ‘communication of data’ criteria. He was able to think of other ways that the data could be presented, analysed, measured or re-stratified.

Table 10.2 Question Data for NORS

NORS	Prob Valid	Measurement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	0		2	0	2	0	0	0		0
ERROR	0	2	0		2	1	2	2		0		2
PRS1|2	0	0	0	2	1	2	2	0	1	2	1	1
PRS4	0	0	1	2	2	1		2	2	0	0	2
FIT												
DIE					1	0	0					0
MEAN	0	0.5	0.3	2	1.6	0.8	1.5	1	1	0.5	0.5	1

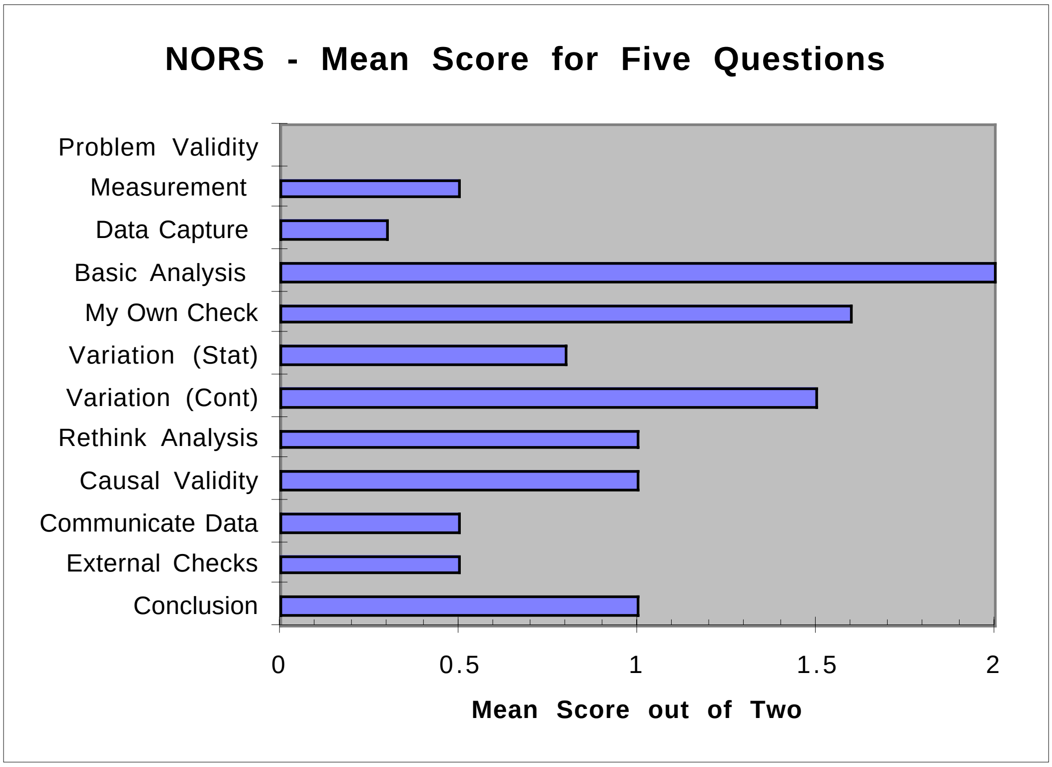

Figure 10.8 Criteria Responses for NORS
[Note the 6 criteria in which NORS scored 1 or more.]

MORTA

MORTA is a female student, in the 17-19 year age group. Her highest school level mathematics studied was Form 7 'Mathematics with Statistics'. She was enrolled in the Stage 1 statistics paper at the time of her participation in the study.

MORTA has a mean score: 1 or greater for 2 criteria; between 0.5 and 1 for 2 criteria; between 0.1 and 0.5 for 5 criteria and no response in 3 criteria (see Table 10.3 and Fig 10.9). These mean scores give MORTA an informal rank of 4= (out of six people).

Only in two questions did MORTA respond with ideas about statistical variation but these were the questions in which all the students recognised rudimentary variation. There was no other indication that statistical variation is an inherent type of thinking for her, and she gave very limited responses in the ‘transnumeration’ criteria.

Table 10.3 Question Data for MORTA

MORTA	Prob Valid	Measurement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	1		2	0	2	0	0	0		1
ERROR	0	0	0		2	1	0	0		0		0
PRS1|2	0	1	0	0	0	1	2	1	0	0	0	0
PRS4	0	0	0	0		0	2	0	1	0	0	1
FIT	2	0	0	0	1	0	1	0	2	0	0	1
DIE					2	0	1					0
MEAN	0.4	0.2	0.2	0	1.4	0.3	1.3	0.2	0.8	0	0	0.5

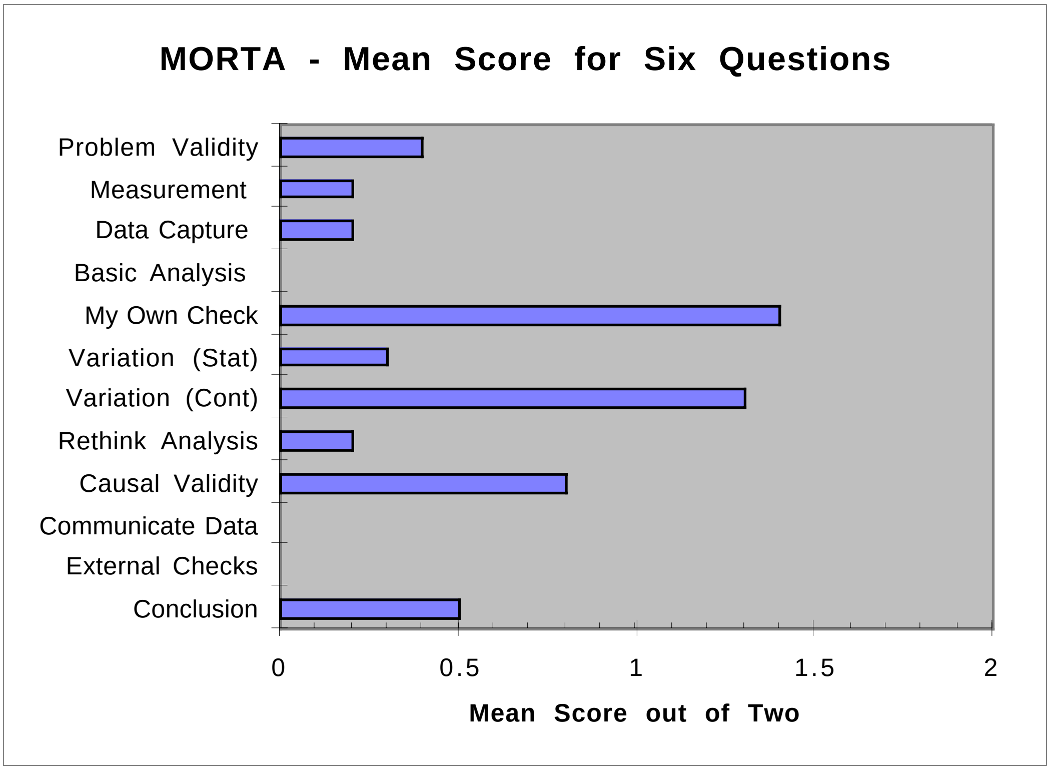

Figure 10.9 Criteria Responses for MORTA
[Note the 2 criteria in which MORTA scored 1 or more.]

EAGLE

EAGLE is a male student from a professional science background, in the 40+ year age group. His highest level of statistics studied was Stage 2.

EAGLE has a mean score: 1 or greater for 2 criteria; between 0.5 and 1 for 4 criteria; between 0.1 and 0.5 for 2 criteria and no response for 4 criteria (see Table 10.4 and Fig 10.10). These mean scores give EAGLE an informal rank of 3 (out of six people).

The 'prison newspaper questions 1 & 2' stimulated some well formed ideas about statistical variation, possibly because of the visualisation of a time series plot from the text. He is the only student to score at least 1 in this criterion. What is surprising is that EAGLE believed the die was biased in the 'die question', gave individual and system 'context explanations for the variation' and gave no consideration to 'statistical variation'. None of the criteria for 'transnumeration' are triggered for EAGLE.

Table 10.4 Question Data for EAGLE

EAGLE	Prob Valid	Measurement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	2		1	1	2	0	0	0		1
ERROR	0	0	0		2	1	1	0		0		0
PRS1|2	0	0	0	0	1	2	1	0	1	0	0	1
PRS4	0	0	0	0		1	1	0	0	0	1	0
FIT	2	0	2	0	0	1	2	0	2	0	0	1
DIE					0	0	2					0
MEAN	0.4	0	0.8	0	0.8	1	1.5	0	0.8	0	0.3	0.5

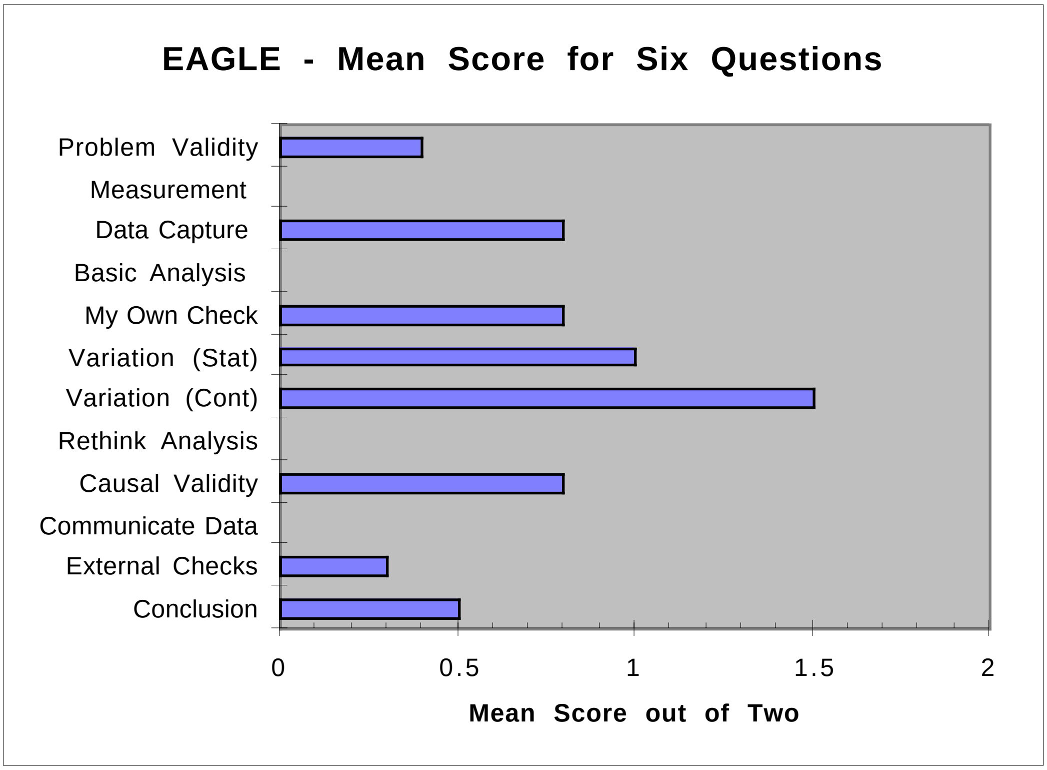

Figure 10.10 Criteria Responses for EAGLE
[Note the 2 criteria in which EAGLE scored 1 or more.]

ISA

ISA is a male student from a semi-skilled background, in the 20-24 year age group. His highest school level mathematics studied was Form 5 Mathematics. He was enrolled in the Stage 1 statistics paper at the time of his participation in the study.

ISA has a mean score: 1 or greater for 2 criteria; between 0.5 and 1 for 2 criteria; between 0.1 and 0.5 for 4 criteria and no response in 4 criteria (see Table 10.5 and Fig 10.11). These mean scores give ISA an informal rank of 4= (out of six people).

The 'prison newspaper questions 1 & 2' did not stimulate ideas about statistical variation. However in the 'fitness newspaper question' he gave a good description of statistical variation. His visualisation and explanation was for a time series plot in the same manner as NORS and EAGLE. Only one criterion of 'transnumeration' is triggered in one article for ISA.

Table 10.5 Question Data for ISA

ISA	Prob Valid	Measu rement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	0		1	0	2	0	0	0		0
ERROR	0	0	0		2	1	1	0		0		0
PRS1|2	0	0	0	0	0	0	2	0	0	0	0	0
PRS4	0	0	0	2	1	0	1	0	0	0	0	0
FIT	0	0	0	0	1	2	2	1	1	0	1	1
DIE					2	0	0					1
MEAN	0	0	0	0.7	1.2	0.5	1.3	0.2	0.3	0	0.3	0.3

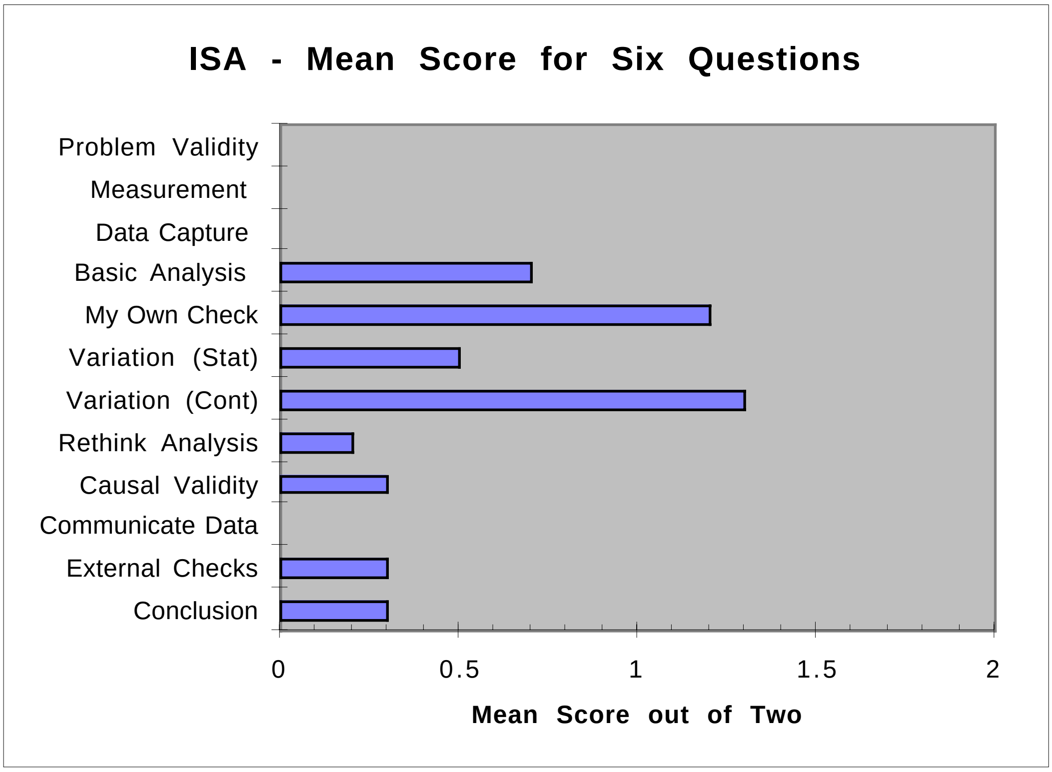

Figure 10.11 Criteria Responses for ISA
[Note the 2 criteria in which ISA scored 1 or more.]

TEP

TEP is a male student from a technician background, in the 30-39 year age group. His highest school level mathematics studied was Form 5 Mathematics. He had never enrolled in a statistics course yet, through the course of his employment as an agricultural technician, he had used statistics, including experimental design, statistical summaries and significance tests. He was also using statistics in other subjects studied at university.

TEP has a mean score: 1 or greater for 3 criteria; between 0.5 and 1 for 1 criterion; between 0.1 and 0.5 for 3 criteria and no response for 5 criteria (see Table 10.6 and Fig 10.12). These mean scores give TEP an informal rank of 4= (out of six people).

Only in two questions did TEP respond with ideas about statistical variation but these were the questions in which all the students recognised rudimentary variation. There is no other indication that statistical variation is an inherent type of thinking for him. Only one criterion of 'transnumeration' is triggered in one article.

Table 10.6 Question Data for TEP

TEP	Prob Valid	Measurement	Data Capt	Basic Analy	Own Check	Variat (Stat)	Variat (Cont)	Rethnk Analy	Causal Valid	Comm Data	Ext Check	Corr Concl
MAP	0	0	0		2	0	2	0	0	0		0
ERROR	0	0	0		1	1	1	0		0		0
PRS1|2	0	0	0	0	1	1	2	1	1	0	0	1
PRS4	0	0	0	2	1	0	2	0	0	0	0	0
FIT												
DIE					1	0	0					2
MEAN	0	0	0	1	1.2	0.4	1.4	0.3	0.3	0	0	0.6

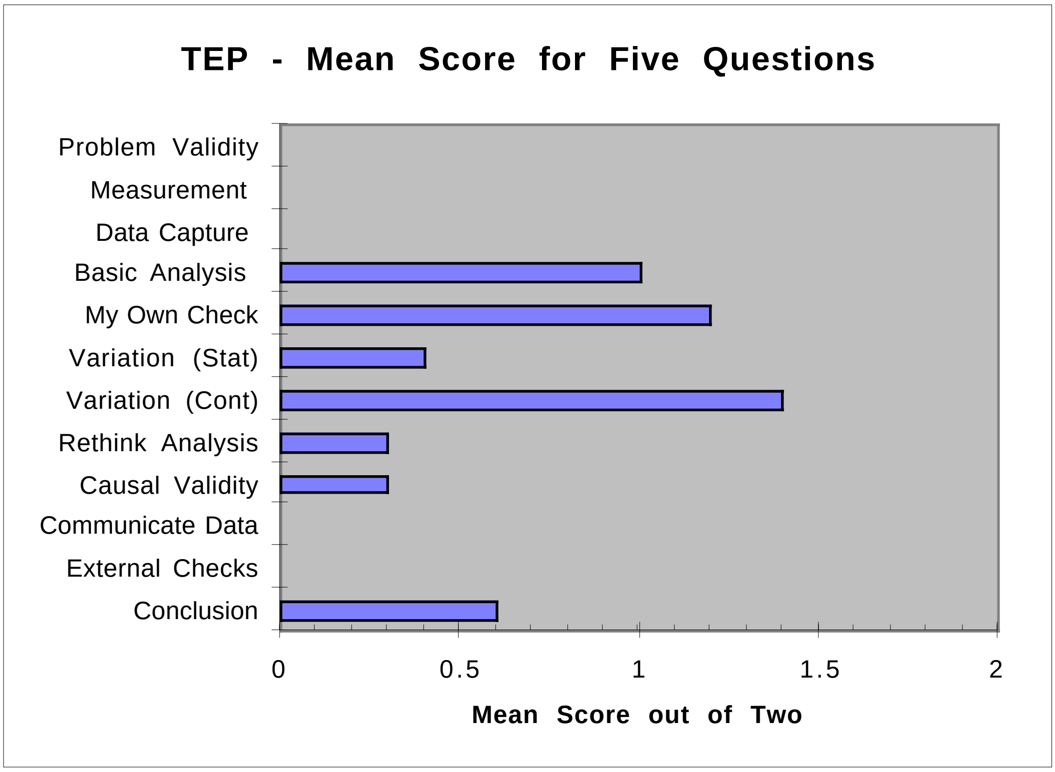

Figure 10.12 Criteria Responses for TEP
[Note the 3 criteria in which TEP scored 1 or more.]

10.6 Discussion

Analysis of Questions

Evaluating statistically based information is not part of the University of Auckland Stage 1 statistics course offered to students. The Stage 1 statistics course spends one week discussing the investigative cycle, polls and surveys, and experimental and observational studies. JOY, MORTA and ISA should have carried out small scale statistical investigations at school. TEP and EAGLE, through the course of their work, would have been involved in statistical or scientific investigations. Despite the fact that most of the students may have experienced the process of an investigation they did not use the stages of the investigative cycle to interpret and evaluate the information. This suggests that the students need a thinking tool or framework to aid a systematic and fairly comprehensive evaluation, and judgement, of statistically based information. It also suggests that students need instruction and practise in evaluating articles, as the responses to the articles tended to be reactions to the information presented.

The analysis part of the cycle is the part that had the most reactions from the students, which is not surprising given that this a feature of the statistics course. The first parts of the investigative cycle such as 'problem validity', 'data capture validity' and 'measurement validity', may need to be addressed in statistics courses. If a graph is presented, students' attention is focussed on the graph rather than the text. My conjecture is that instruction on media articles in the statistics course has focussed on misleading graphs rather than on evaluating the claims made in the text of the articles. A graph may be misleading but this does not imply that the study and resultant claims are suspect. Valid information can sometimes be extracted from bad graphs. There may be a need to address this issue in instruction.

Statistical variation, and, in particular, statistical variation in small samples, is not present as an underlying idea or principle that interconnects the statistical knowledge base. It does not appear to be an inherent part of all statistical information and thinking. The statistical variation ideas that are present, are centred around a time series plot with a trend rather than sampling distributions and their relationship to theoretical population models. For example the students did not seem to have visual imagery for variation in boxplots in relation to the sample size and the population boxplot, nor an image of a Normal distribution and distributions of samples with different sample sizes from that population.

The ‘transnumeration’ type of thinking, which requires a production of knowledge related to data and context (e.g. thinking of ways that data could be interpreted or represented), does not seem to be part of the students’ repertoire. ‘Reasoning with models’ that are given, or easy to produce visually from the given data, such as a time series plot, seem much easier for students than recognising and making their own constructions of applicable statistical models. Producing context explanations for all questions seemed natural to the students, except for inconsistency in producing individual-based and system-based explanations. These, too, require a production of knowledge (e.g thinking of ways that the context could be interpreted) but do not involve data, and therein may lie the difference. Students are taught to reason with given statistical models but they probably do not have much experience *determining the models with which to reason* or thinking of other statistical ways to reason about a situation (‘transnumeration’).

The reaction type of thinking used by the students is mostly critical, rather than affirmative, of information. In evaluating an article students must weigh up what information they are willing to believe or not believe and then give reasons for their judgement. From my informal observations of my own students in the classroom, when trialing some findings of the second exploratory study in teaching, the critiquing of an article is not easy. The students needed to practise this skill. Therefore I would conclude that students need some scaffolding for reading, understanding and evaluating a statistically based report.

Analysis of Individual Students

There are no discernible patterns in the individual student responses. Whenever I conjecture a possibility there is an anomaly. The two students with the least appreciation of statistical variation are MORTA, who had a good background in statistics, and TEP who had no formal background in statistics. The two students, JOY and NORS, who had the best scores in 'transnumeration' and overall had the most categories stimulated, are in the oldest and youngest age group categories and had different school and world experience backgrounds. An incorrect conclusion in the 'die question' was given by EAGLE, NORS and MORTA, the first two of whom would probably not have had die tossing experience at school, whereas MORTA has a school background which should have included such experimentation. TEP and JOY gave correct conclusions to the 'die question' yet their backgrounds are fundamentally different.

From the data collected, an informal ranking of students on the development of their statistical thinking would be NORS(1), JOY(2), EAGLE(3) and MORTA, ISA and TEP(ranked equally on 4). On consideration of their statistical backgrounds and world experience no conjectures are possible.

Summary

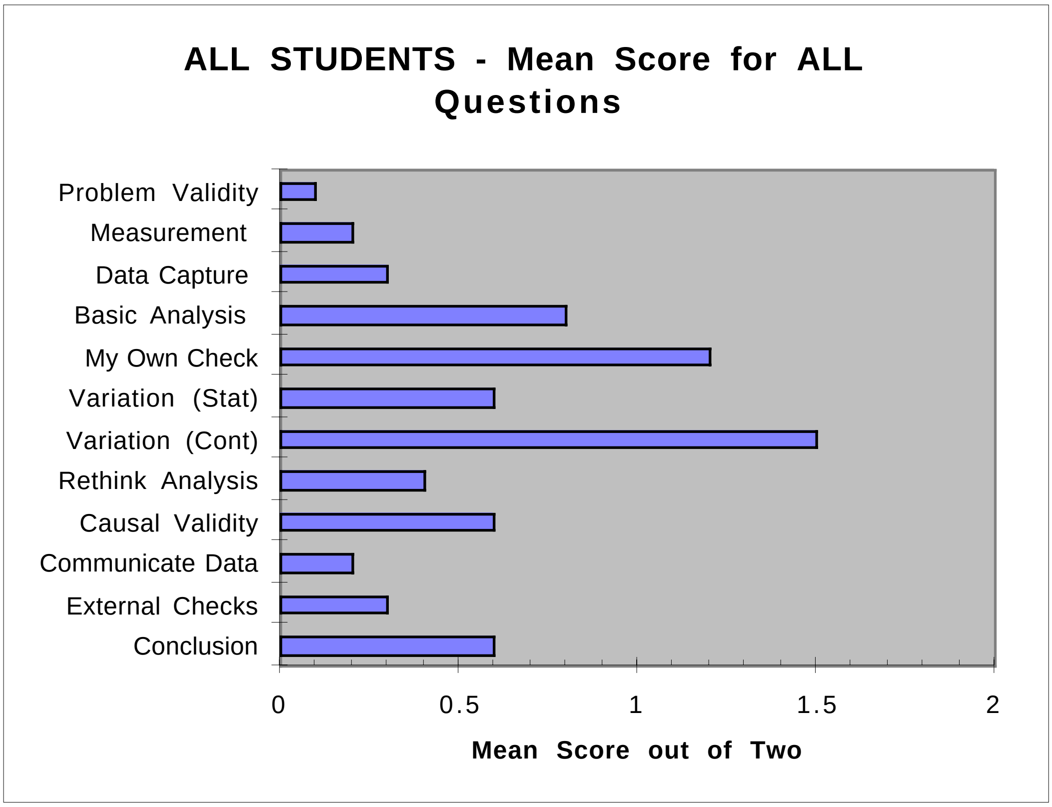

Figure 10.13 Student Response Data for Judgement Criteria
[Note: First stages of investigative cycle are not triggered as much as the analysis stage.]

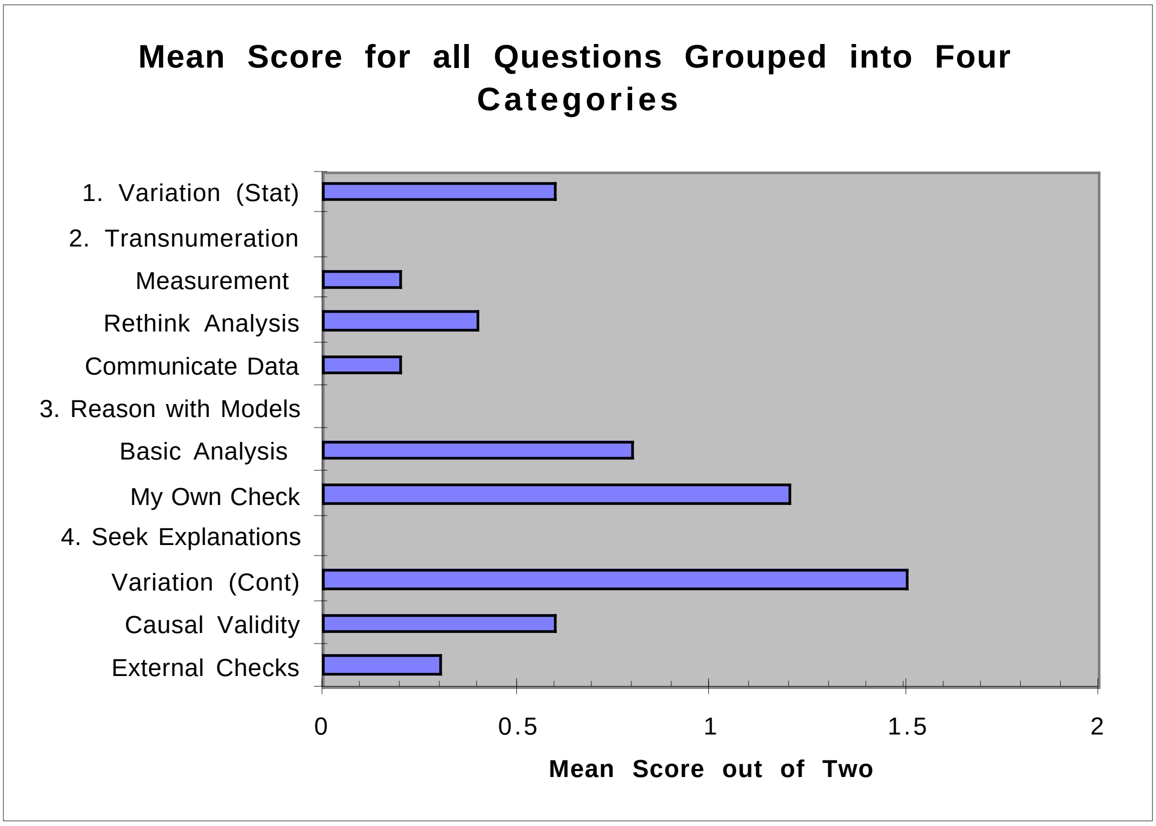

Figure 10.14 Student Response Data Grouped into Four Categories
[Note: 'Reasoning with Models' and 'Seeking Explanations', or thinking with context, are triggered more than 'Transnumeration' and 'Statistical Variation', or thinking with data and context.]

The judgement criteria or "critical list" do not provide information about which particular questions in each criterion are not being prompted in students, but they do provide information about areas of statistical thinking that need further development. Therefore, unlike Gal et al. (1995) who perhaps were thinking of one short "critical list", we believe that our categorised judgement criteria or "critical list" will inform teaching and learning, not only in the evaluation of media articles, but also in the reporting of statistically based information and in the carrying out of statistical investigations.

From the interviews it appeared that all the students had a critical attitude particularly when their responses to 'seeking explanations' from the context are considered. This innate critical attitude did not seem to be triggered consistently because, firstly, they did not have a connected framework to use to respond to such situations and secondly, the students did not easily think, or know how to think, with data, or with data and context simultaneously (see Figs. 10.13 and 10.14). This suggests that a framework or thinking tool or judgement criteria would help students to construct considered responses to information. It would also help to combat the tendency of students to react only to information that is present since it would alert them to information that is not present.

Considering the questions given to the students were not the usual type of questions given in their statistics courses, it could be assumed that the current teaching approach, course content and assessment are not promoting the statistical thinking we would like to develop in students. Therefore the statistical thinking framework developed in Chapter 9, from which the judgement criteria were explicated for this analysis, may be useful for developing teaching approaches which cultivate such thinking.