MATHEMATICS TEACHING-RESEARCH JOURNAL ONLINE
VOL 8, N 1-2
Fall and Winter 2015/16

Teaching and Assessment of Statistics to Employees in the NZ State Sector

Murray Black

Auckland University of Technology, New Zealand
murray.black@aut.ac.nz

Abstract

This paper covers my experiences and reflections both in the teaching and the assessment of statistics to state sector employees. These employees undergo courses in statistics where the taught content and assessments are based on unit standards which have been designed to be relevant with the professional work of the participants. These unit standards form part of the newly-developed National Certificate in Public Sector Services (Official Statistics). This paper uses an action research methodology to ascertain what improvements could be made in the preparation and teaching of the statistical content along with the design and implementation of these unit standard assessment tools.

A review is carried out of the teaching based on the performance criteria of the unit standards, using the first cohort, with the intent of making improvements to materials, teaching approaches and tutorial support for the following cohorts of learners. Also for the learner, support systems recommended include new learning materials, mentoring systems, course organisation changes and tutorials.

An examination is carried out of the questions with respect to their type according to a statistical reasoning level scale and order of presentation within each tool. From the first cohort of students, observations of learner's response times involving dates of requests, submission and re-sits are taken from an extant database. From both a summary of the learner's barriers to completion and an analysis of response times a series of changes for when the tools are to be used again are recommended. These changes involved a redesign of the assessment questions.

Keywords: *Unit Standard, Statistical Reasoning, Perfor*mance Criteria

MATHEMATICS TEACHING-RESEARCH JOURNAL ONLINE
VOL 8, N 1-2
Fall and Winter 2015/16

1 Introduction

The Education Act 1989 established the New Zealand Qualifications Authority (NZQA) to, "oversee the setting of standards for qualifications in secondary schools and in post-school education and training"(NZQA, 1989, section 253(a)).Also it set the legislative base for the National Qualifications Framework (NQF) as one in which

"…All qualifications... have a purpose and a relationship to each other that students and the public can understand and there is a flexible system for the gaining of qualifications with recognition of competency already achieved" (NZQA, 1989, section 253(c)). Assessment in the units of learning (unit standards) registered on this framework focused on the measurement of learner performance against published standards called performance criteria (New Zealand Qualifications Authority, 1991 cited in National Qualifications Project Team 2005). Assessment questions are designed to measure achievement with respect to each question being designed to satisfy a particular performance criterion within each unit standard.

In the second half of 2007, a Certificate in Public Sector Services in Official Statistics was introduced as a qualification designed for people working in the State Sector. This qualification provides learners with the range of vocational knowledge and skills required to be able to read, identify and interpret statistics having a state sector context. This certificate consists of 40 credits of unit standards with 24 credits being allocated to those standards which involve applications of statistics. These 24 credits are made up by four core unit standards which are assessed separately. These assessments are based on a series of performance criteria within each unit standard. The unit standards are:

- US 23268: Level 4, 8 credits. Interpret statistical information to form conclusions for projects in a state sector context.
- US 23269: Level 5, 8 credits. Evaluate and use statistical information to make policy recommendations in a state sector context.

- US 23270: Level 4, 4 credits Assess a sample survey and evaluate inferences in a state sector context.
- US 23271: Level 5, 4 credits Resolve ethical and legal issues in the collection and use of data in a state sector context.

A report involving official statistics is selected beforehand and used as an exemplar for the assessments. To obtain credit for each standard, the learner is required to answer all the questions pertaining to the selected report for the assessment correctly. In the case of an answer being incorrect or not fully answered, the learner is offered a re-sit where they review their answer(s) and resubmit (http://www.statisphere.govt.nz/certificate-of-official-statistics/default.htm).

.

1.1 Purpose of Questions

The key objective of the questions is to put the learners in the position of having to read and interpret reports over a range of statistical concepts. In doing so they need to bear in mind the overall objective of the report and how the statistics within the various reports informed the various answers to policy questions. Questions incorporated the following areas;

- assessing the relevance of data
- finding and selecting data relevant to a policy question
- interpreting findings
- making policy recommendations based on the data
- explaining how a particular piece of statistics could be performed with a possible result in the context of the report. For instance, where you would calculate a confidence interval for the difference between means and how the results would be interpreted

- interpreting possible results and designing a data collection to answer a policy or research question and
- explaining the limitations of a chosen exemplar, for instance stating the omission of a margin of error in the report along with possible consequences.

In the case of unit standards 23268 and 23270 a list of topics was developed then questions were written around those topics to ensure full coverage. For unit standards 23269 and 23271 the questions were written around each performance criteria. The assessment questions are a critical aspect of a unit standard as they measure learner performances against a set of learning objectives laid down in the unit standard documents as performance criteria.

1.2 Framework Mappings

A domain of educational activity that can be associated with the requirements of these four assessment tools is the cognitive domain as identified by Bloom (1956). This domain involves knowledge and the development of skills and has six major categories (Bloom 1956). Consequently Bloom's taxonomy was updated and the update published in 2001 by Anderson and Krathwohl. From their update new categories were described as follows:

1. **Remembering**: Retrieving, recognising and recalling relevant knowledge from long term memory.
2. **Understanding**: Constructing meaning from….written……messages through interpreting, classifying, summarizing, inferring, comparing, and explaining.
3. **Applying**: Carrying out or using a procedure through executing, or implementing.
4. **Analysing**: Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.

5. **Evaluating**: Making judgments based on criteria and standards through checking and critiquing.
6. **Creating**: Putting elements together to form a coherent of functional whole; reorganizing elements into a new pattern of structure through generating, planning, or producing.

(Anderson & Krathwohl, 2001, pp. 67-68)

We wish to map from the six category levels as defined by (Bloom 1956), denoting degrees of difficulty alongside their corresponding six revised categories as defined by (Anderson & Krathwohl 2001). Then these six categories are mapped alongside, with category levels 5 and 6 combined, to the instructional domains as defined by (DelMas 2002) which pertain to descriptions of tasks and then alongside into five levels of statistical reasoning as proposed by Garfield (2002).

These mappings in hierarchical level order are shown in Table 1 below:

Level	Bloom's Taxonomy Objective	Revised Taxonomy Anderson & Krathwohl	Instructional Domains Teaching	Reasoning Framework Assessment
1	Knowledge Recall	Remembering Recall	Literacy Identify	Idiosyncratic Knows
2	Comprehension Meaning	Understanding Meaning	Literacy Describes	Verbal Defines
3	Application Context	Applying Context	Reasoning Why?	Transitional Partial Understanding
4	Analysis Distinguishes	Analyzing Organizing	Reasoning How?	Procedure Application
5	Synthesis Contextual Links	Evaluating Criquing	Thinking Apply	Integrated Process Complete Understanding

6	Evaluating Construct Report	Creating Reorganizing		

Table 1
Mapping Links from Taxonomies to Statistical Reasoning Levels

A recent theory of learning which has been widely accepted in education communities stems from earlier work by Jean Piaget, and has been labelled as "constructivism" (Garfield, 1995). This theory describes learning as actively constructing one's own knowledge (Von Glasersfeld, 1987). As the level of statistical reasoning required to solve a problem increases, one has to construct more knowledge and display higher levels of learning, according to the taxonomies in Table 1, of statistical concepts in order to succeed. This makes the required statistical concepts at these higher levels more difficult than others at the lower levels.

The relevance of these five statistical reasoning levels, to the successful completion of an assessment, is such that the correct statistical reasoning required to either read and understand or compose a report should follow all of these levels as steps. Firstly one needs to know the relevant statistical concepts and what they mean. Secondly these concepts would need to be applied to the context of the report and be integrated where relevant to other parts of the report. Finally appropriate conclusions need to be drawn based on these statistical concepts in order to meet the objective of the report. Although not originally designed to be such, ideally the order of questions in a typical assessment of a unit standard should follow these reasoning steps and cover all levels.

2 Aim

This paper considers the following research question: how can the effectiveness of these four unit standard assessment tools are improved in terms of their structure and implementation? To answer the question this paper sets out to evaluate whether the current unit standard assessment tools have been designed appropriately. Also an analysis of the learner responses to these assessments is reported. The objective of this evaluation is to establish a basis for considering changes to the assessment following the findings from the first pilot cohort of learners. As a result of this evaluation by using an action research methodology, this paper describes how both the structure and operation of these unit standard assessments were reviewed and changed where justified by the research for the next cohort of learners.

3 Methods

3.1 Methodology

The research methodology followed here is action research as it's a form of collective self-reflective enquiry undertaken by participants in order to improve the rationality of their own educational practice (Kemmis & McTaggart 1988). The practice being improved here is the design and implementation of these unit standard assessment tools. According to Kemmis and McTaggart (1988) an action research study improves education by changing it and learning from the consequences. Here data pertaining to the first cohort of workplace learners was collected. Then changes were made to the assessment tools in readiness for the second cohort in 2008. The research process consisted of one action research cycle (Kemmis& McTaggart 1988 pp14) After the end of the assessment cycle; the assessments were reviewed with the intent of making improvements. The research design involved two main sources of data:

- an examination of the original assessment questions and

- an analysis of the learner responses to these assessments.

3.2 Examination of Assessment Questions

The data reviewed was already available in Learning State's Assessment Guides for each unit standard. An examination of the 71 questions across the four standards was carried out with respect to the type and order of questions within each standard. A classification of each question into its appropriate statistical reasoning category (refer Table 1) was performed even though the questions were not designed that way.

3.3 Learner Responses to these Assessments

The data reviewed was part of an "extant data base" already available in learning state records and an examination was carried out of the spreadsheet which gave submission, re-sit submission and request dates for each set of assessment questions. Learners requested their assessments on a recorded date then the date of receipt was recorded. The reply date to the learner with any re-sit questions was recorded along with the submission date of their answers. The pass date was also recorded.

The participants in the first cohort give informed consent to the use of their re-sit assessment answers after their passes had been recorded and submitted. A through analysis was performed on the re-sits in terms of questions that needed to be re-sat, issues associated with question design, teaching and assessment along with an analysis of timing. The number of successful completions which were 13, 11, 12 and 10 respectively from the four standards, 23268 to 23271 inclusive, were analysed with respect to timing, learner's answers to incorrect questions and barriers to completion for the learners. Also the impact on the level of statistical reasoning required in each question requiring a re-sit was assessed.

4 Results

This section outlines the results that were obtained from an examination of assessment questions and learner responses.

4.1 Type of Questions

Tables 2 and 3 (refer appendix) categorize the answers provided for each question from the assessor's point of view. Data using a scale of 1 to 5 (DelMas 2002) depending on the level of reasoning (Garfield 2002) was recorded. These levels correspond to those in Table 1.

1 = Identify (pick out answer from the report).
2 = Describe with no requirement to answer in context
3 = Why this? – answer requires context to the report in specified part only
4 = How? – some link to more than one part of report is required in context
5 = Apply – all links to relevant parts of report are required in context.

So a reasoning category in the range 3 to 5 inclusive would indicate that the question requires some linkage of the statistical analysis to the context of the report. From tables 2 and 3 we analyse both the order and frequency of the levels of statistical reasoning required to answer the questions in section 4.2.

4.2 Order of Questions

Originally the orders of the assessment questions were designed so the content was tested in the order it appeared in each of the performance criteria pertaining to each unit standard. Table 4 below gives a cross-tabulation between the required level of statistical reasoning to answer the question and each unit standard with the column totals giving the total number of questions within each unit standard assessment tool.

Level	US 68	US 69	US 70	US 71
1	4	2	2	0

2	5	0	9	3
3	6	6	6	9
4	3	5	1	2
5	0	5	1	2
Total:	18	18	19	16

Table 4 Number of Questions in each Level Category by Unit Standard

Combining the levels 1&2 and levels 4&5 and using percentages so columns total 100% we get Table 5 below:

Levels	US 68	US 69	US 70	US 71
1&2	50%	11%	58%	19%
3	33%	33%	32%	56%
4&5	17%	56%	10%	25%
Total:	100%	100%	100%	100%

Table 5
Percentages of the Questions of each Unit Standard

We note that the two level 4 standards, US 23268 and US 23270 have considerably more questions requiring the lowest two levels of statistical reasoning to answer than US 23269 and US 23271.This confirms what we would expect with the two level 4 standards. Conversely US 23269 have over half its questions requiring statistical reasoning to levels 4&5, at 56%. US 23271 have a lower percentage in the higher level category than expected and 56% in the middle category. This represents contextual links at the lowest level. Ideally these two percentages should be reversed. This could be due to some of its questions revisiting basic concepts as part of scaffolding into questions involving reasoning to a higher level, assuming the validity of Bloom's Taxonomy, throughout the assessment.

If we run a chi-square test using the data tabulated in Table 5 at the 5% level of significance we would conclude that level and unit standard are significantly related ($\chi^2 = 17.7 > 12.592$ at $\alpha = 0.05$ and $\nu = 6$). So irrespective of how the questions were

designed we have significant differences in the proportions of questions containing the various levels of reasoning over these four unit standards.

Figure 1 below displays the levels of reasoning by question order for each unit standard where the line graphs follow the order of questions in the assessment. The continuous nature of these graphs reflects the process over time taken to work through these assessments in question order from low to high.

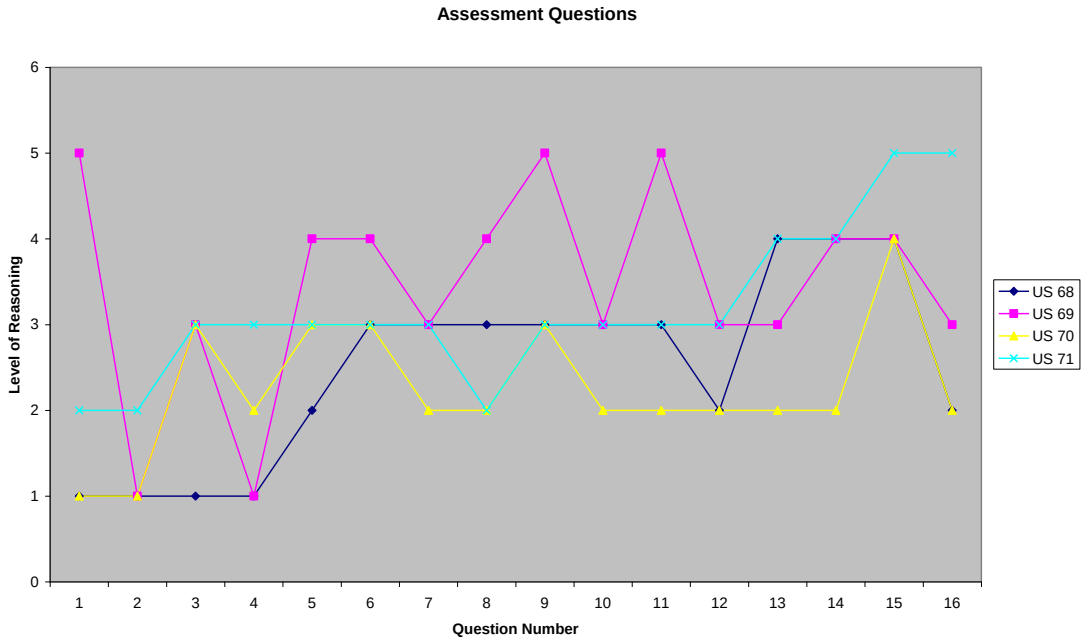

Figure 1
Levels of Statistical Reasoning by Question Order

From Table 1 and Figure 1 we observe that the order of questions for assessing US 23271 follows the levels of statistical reasoning upwards, low to high, with only one drop between questions 7 and 8, by starting off at level 2 and rising to finish at level 5. Also the upward trend in statistical reasoning levels is shown by US 23268 for questions 1 to 15 and in the main by US 23270 except for the last two questions. US 23269 appears to not follow the trend of the other standards and this would be largely due to its questions being designed to sit separately under each of the elements 1 to 4.

4.3 Re-Sit Analysis

In most cases follow up re-sit questions were done by email. Some were done verbally so for these there is no paper trail. There were face to face meetings when the author was down at Statistics New Zealand and also there were phone conversations. Usually any outstanding issues with re-sit questions were cleared up reasonably quickly. On the whole candidates answered most questions correctly first time however there was a high percentage of learners needed a re-sit on at least one question in every unit standard, before passing, as Table 6 shows:

Percentage of Candidates requiring a Re-sit before passing	
US 23268	77%
US 23269	67%
US23270	50%
US23271	60%

Table 6
Re-sit Percentages

Out of those who requested the assessment, Table 7 shows percentages of candidates that passed first time:

Percentage of Candidates who passed first time	
US 23268	23%
US 23269	31%
US23270	50%
US23271	33%

Table 7
First Time Passing Percentages

The common issues involving teaching, assessment design and candidate issues that may have influenced re-sits on at least one question are identified in sections 4.3.1 to 4.3.3 below:

4.3.1 Teaching Issues

There was a lack of teaching coverage of some concepts. Two examples of these noted knew what cyclical variation was in a time series (US 23268) and what performance indicators represented. (US 23268 Q15). In addition, learners had more difficulties with the two quantitative assessments as opposed to the two qualitative assessments.

4.3.2 Assessment Design Issues

Learners were unable to find examples of required content to answer a particular question in the report so it could be answered in context. One area of difficulty was finding an example of a bivariate analysis (US 23268 Q10). There were difficulties in explaining concepts not covered in report. For example, confidence intervals for differences between proportions. This was largely due to assessment requirements where all the concepts in a learner's answer had to be related to a maximum of two reports for each assessment. (US 23270 Q13). Also overall questions in assessment were not being clear enough about what was required to answer the question.

4.3.3 Learner Assessment Issues

In several instances, questions not answered completely. An example of this was stating fully all the 12 privacy principles or only partly answering a question and another

example was a failure to provide appropriate context when required. In other cases the answers were too brief and not enough detail given e.g. recommendations to management. A common occurrence was that only one part of question was answered. For example, the learners were able to distinguish between the two sampling methods, stratification and clustering but were unable to explain why stratification was preferred in the report (US 23270 Q17).

4.3.4 Types of Errors

Overall re-sits occurred because learners were incorrectly explaining statistical concepts in answering questions within the level 4 units whereas re-sits were largely due to incomplete answers to questions in the level 5 units. Part of the reason for this was that the question was not making it clear to candidates what was required for a full answer to the question. All questions and re-sits were handled by email. Specific questions (refer appendix) within each unit standard that required a re-sit are shown in Table 8 below:

Question Number	23268	23269	23270	23271
1	0	2	0	3
2	0	0	0	0
3	1	1	1	0
4	0	0	0	0
5	1	1	1	0
6	2	0	2	0
7	1	1	1	1
8	1	1	4	0
9	0	0	0	0
10	4	4	2	0
11	0	1	2	2
12	1	0	0	0
13	5	2	1	0
14	0	1	1	0
15	3	0	1	0
16	2	2	2	3
17	1	4	4	
18	2	0	0	

19			3	

Table 8
Frequency of Questions needing to be Re-Sat by Unit Standard

As an overall summary, the nature of the unsatisfactory responses described in three categories of error classification made by the candidates within each unit standard requiring re-sits are shown in Table 9.

Error Classification	US 23268	US 23269	US 23270	US 23271
Incorrect	9	2	7	2
Correct with no context	0	2	8	2
Partially correct with/without context	15	16	10	5

Table 9
Error Frequency

We observe that the major source of error, over all the standards, was that questions were not answered fully enough. Those questions involving at least three learners having to re-sit had the following characteristics shown in table 10.

Unit Standard	Topic	Characteristic
23268	Bi-variate	No practical example
	Time Series Cyclical Variation Irregular Variation	 Lack of knowledge Lack of knowledge
23270	Statistical Measures	No proper conclusion
	Weighting of Observations	Lack of knowledge
	Sampling Impacts Stratification	 Not attempted

	Clustering	Not attempted
23269	Data Collection Elements	Unclear
23271	Principles of Privacy Act	Not complete
	Data Collection Issues	Recommendations not complete

Table 10
Characteristics of Re-sit Questions

A chi-square test was carried out on the data to see if error classification and level of unit standard were significantly related (Table 11).

Error	Level 4	Level 5
Incorrect	16	4
Partially Correct	33	25

Table 11
Error Frequencies by Level

The question response has been recorded as "not complete" if no context had been provided so the frequencies are sufficiently high enough to validate this test.

We conclude that error classification and error are not significantly related as χ^2 = 3.33 < 3.841 at the 5% level of significance.

4.3.5 Links of Re-sit Questions to Statistical Reasoning Levels

If we tabulate the mean number of re-sit questions per candidate out of the candidates that needed to re-sit and put alongside the number of re-sit questions in each statistical reasoning level, we obtain the results shown in Table 12.

Unit Standard	Mean	Statistical Reasoning Level				
		1	2	3	4	5

23268	2.4	1	7	8	8	0
23269	2.5	0	0	10	3	7
23270	4.2	0	13	7	1	4
23271	1.5	0	3	3	0	3

Table 12
Mean Number and Frequency of Re-sit Questions

We observe in Table 13 that if we rank the mean next to the median of the reasoning levels, taken as a measure of the complexity of a question, from low to high over the four standards we get:

Rank Order	Mean Number of Re-Sit Questions	Median Reasoning Level
1	23271	23270
2	23268	23271
3	23269	23268
4	23270	23269

Table 13
Rank Orders of Unit Standards

The main observation from Table 13 is that US 23270 has the highest mean number of re-sit questions but the lowest median of the statistical reasoning levels when all the others approximately matched differing by only one place. This would suggest that there were difficulties in describing some of the basic concepts having a quantitative nature which were prevalent in US 23270. We wish to examine the pattern of the levels of statistical reasoning required in the 78 questions in total that needed to be re-sat by the learners. This pattern was compared with the overall levels of statistical reasoning required to answer all the questions.

Statistical Reasoning Level	1	2	3	4	5
Total Number of Re-sit Questions	1(1%)	23(29%)	28(36%)	12(15%)	14(18%)
Total Number of					

Questions in Assessment	8(11%)	17(24%)	27(38%)	11(15%)	8(11%)

Table 14
Number of Questions by Statistical Reasoning Level

The main observation in Table 14 is that we have a higher proportion than expected of re-sit questions in the reasoning categories of 2 & 5. This highlights difficulties the learners are having in the areas of explaining statistical concepts and giving their explanations in context to an objective of a report. The reasons for this could be one or more of the following:

1. assessment questions not aligned to statistical reasoning levels
2. not enough time provided for learners to grasp concepts.
3. not enough examples provided in the presentations, and
4. lack of suitable reports for assessment purposes.

A chi-square test was carried out to see if there are there any significant differences in the proportion of questions that needed to be re-sat requiring each statistical reasoning category compared to the overall proportion of questions requiring each statistical reasoning category using the data in Table 13. We get: $\chi^2 = 11.09 > 9.488$ at $\nu = 5\text{-}1 = 4$ degrees of freedom at $\alpha = 0.05$.

So we have significant differences in the proportions with higher proportions of the re-sit questions requiring levels 2 & 3 categories of statistical reasoning and lower proportions of re-sit questions being in the levels 1, 4 & 5 categories of statistical reasoning.

Over the re-sit questions a chi-square test was carried out to see if the level of the unit standard assessment tool and the required statistical reasoning level were related.

The four unit standards were collapsed into two categories in Table 15 depending on whether they were at level 4 or level 5. This needed to be done as there were too many

expected frequencies less than 5 as only a maximum of 10% of the cells can have expected frequencies less than five, for the chi-square test to be valid.

Levels		
Statistical Reasoning	4	5
1 & 2	21	3
3	15	13
4& 5	13	13

Table 15
Frequencies of Re-Sit Questions

The outcome of the analysis was $\chi^2 = 9.04 > 5.991$ at $\alpha = 0.05$ so we conclude that level of assessment tool and required statistical reasoning level of questions to be re-sat are significantly related. This was expected as we had established earlier that in the design of the questions, the statistical reasoning level required to answer these questions was significantly related to the level of the unit standard.

4.4 Completion Analysis

There was a group that managed to complete each unit standard in time and keep up. Many asked for extensions from the stated three weeks which was not rigidly enforced. There were however some long periods between re-sits and submissions. Table 16 on the next page provides some summary statistics:

Unit Standard	No Sent Out	No of Re-sits	No of Passes	% Pass Rate	Median Completion Time (days)	Range of Completion Times (days)
23268	13	10	13	100	29	7 to 174
23269	13	8	12	92	37.5	8 to 150
23270	12	6	12	100	20.5	5 to 240
23271	12	6	10	83	24.5	4 to 167

Table 16
Summary Statistics for Completion of Unit Standards

From Table 16 the combined median completion time of 65.5 days was greater for the two level 5 unit standards over the two level 4 unit standards at 49 days, which confirms what we would expect with these unit standards being at a higher level.

Help towards completion was provided by timely feedback on re-sit questions to the point, conversations with learners regarding re-sits topics not understood, for example cyclical variation, index numbers and odds ratios were covered with a session with learners face to face. Feedback was provided to the contact at Stats NZ for tutorials.
A summary of the identified key barriers to completion are given in table 17 below.

Completion Barriers	
Design of Assessment	Three weeks completion time
	Too short between presentations
	Fragmentation between the tools
	Absence of clear example
Teacher Dependent	Some concepts not taught fully
	Not enough teaching time
	Difficulties in making up an example
Learner Motivation	Reluctance to request help
	Learners left quantitative assessments till later
	Time management skills
Outside Control	Work commitments of learners
	Effect of instruction diminished over time

Table 17
Barriers to Completion

Overall it was concluded, from table 17 that improving the assessment tools coupled with more learner support with additional teaching would help to overcome these barriers.

Tables 18 and 19 (refer appendix) provide a summary of the days taken to complete each assessment by each learner. From this summary we note the following:

1. The mean re-sit time was substantially higher for the two standards that had the lowest complexity rankings in terms of requirements. In fact for US 23270 and US 23268 approximately one third, 29.9% and 32.8% respectively, of the total learner time spent working on these assessments was working on re-sit questions. Only 2.7% and 2.8% of total time was spent by learners working on re-sits to do with US 23269 and US 23271 respectively.
2. The proportion of re-sits was about the same for each standard with the maximum (0.77) and minimum (0.50) occurring with the two standards 23268 and 23270 respectively of lowest complexity.
3. The % of completions was highest for US 23270 and lowest for US 23269 both in all three categories; one month, two months and three months from delivery.
4. The variability measured by the standard deviation of total times taken for US 23270 is substantially higher at 64.5 days than for the other three standards probably due to the outlier at 240 days.

5 Discussion & Implications

We concluded that the statistical reasoning level and unit standard were significantly related. The level 4 unit standards as expected displayed lower levels of statistical reasoning. Some work needs to be carried out on US 23269 to ensure a more appropriate ordering of questions and minimal work to US 23270. Two key factors causing the need for re-sits were poor knowledge of concepts along with poorly designed questions. Answers requiring re-sits were largely only partially answered. In US 23270 with questions having a statistical reasoning classification of 2 or 5, there were difficulties in explaining statistical concepts and relating them to the overall report objective. The main

barrier to completion was caused by the absence of a report which gave an example of all the statistical concepts being assessed in the standard,

Efforts are underway to build on past experiences of using these assessment tools and to review and improve them by making the following changes:

5.1 Assessment Questions

The questions in unit standard assessments; 23269, 23270 and 23271 were modified to give fewer questions and closer alignment with the five stages of statistical reasoning in question sequences. There were now no overlaps in questions requiring similar answers between the four unit standard assessments. This means we can reduce the number of questions and hence "assessment burden" on the learners and still satisfy the performance criteria. The aim was now to have linkages in assessing components of all four standards across one or two common over-arching reports. Other reports were to be provided on a minimal basis to cover assessed statistical concepts not covered by these over-aching reports. Questions were designed so some optional choice was provided in explaining concepts. For instance, explain in context a confidence interval for a mean or a proportion. Learners would need to determine whether the confidence interval being used in the report was for a mean or a proportion then explain their choice in context. The need for learners to switch reports in order to answer all the questions in the assessments was substantially reduced.

5.2 Seminar Presentations

The order of presentation of the unit standards was changed so now the teaching sequence was 23271, 23268. 23270 and 23269 rather than the order 23270, 23268, 23269 and 23271. This was so we can focus the learners more on the overall objectives of a report along with legal and ethical constraints before we teach the various statistical concepts required. Fine tuning of the production of the worked examples as part of the presentation took place. These showed what was required for a pass for different types of questions. New resource material was produced to assist in the presentations of the unit standard material where there was now more emphasis on performance criteria along with their associated statistical concepts.

5.3 Support System

The backup and mentoring systems for learners were extended. There was the provision of pre-courses in relevant material, after a gaps analysis to ascertain entry knowledge of learners, so learners are better prepared. For example, a pre-course on confidence intervals was planned prior to the delivery of US 23270.

5.4 Future Developments

It is possible to have another cycle of action research to assess the effectiveness of these changes to the design and implementation of the assessment tools. Also changes could be recommended to the unit standards themselves with respect to the performance criteria. These could be rewritten to mirror the sequence of statistical reasoning steps that must occur for the successful application of statistical concepts in the workplace.

Reference

Anderson, L.W., & Krathwohl, D.R. (Eds.). (2001). *Taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*: Complete edition, New York: Longman.

Bloom, B.S. (1956), 'Taxonomy of Educational Objectives: The Classification of Educational Goals', *Handbook I; Cognitive Domain*, New York, USA: Longmans, Green and Co.

DelMas, R.C. (2002), 'Statistical Literacy, Reasoning and Learning: A Commentary', *Statistical Education*, Vol 10(3).

Garfield, J. (1995). How Students learn Statistics, *International Statistical Review*. Vol 63(1). pp.25-34.

Garfield, J. (2002), 'The Challenge of Developing Statistical Reasoning', *Journal of Statistics Education*, Vol10 (3).

http://www.statisphere.govt.nz/certificate-of-official-statistics/default.htm

Kemmis, S., & McTaggart, R. (Eds.). (1988). *The Action Research Planner* Third edition, Deakin University, Victoria.

National Qualifications Framework project team (2005). *The New Zealand Qualifications Framework, Revised Paper*, New Zealand Qualifications Authority, Wellington.

Von Glasersfeld, E. (1987). Learning as a constructive activity. In *Problem of representation in the teaching and learning of mathematics*. C. Janvier (Ed.). pp. 3-17. Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix

US 23268	Level of Reasoning Category	US 23269	Level of Reasoning Category
1	1	1	5
2	1	2	1
3	1	3	3
4	1	4	1
5	2	5	4
6	3	6	4
7	3	7	3
8	3	8	4
9	3	9	5
10	3	10	3
11	3	11	5
12	2	12	3
13	4	13	3
14	4	14	4
15	4	15	4
16	2	16	3
17	2	17	5
18	2	18	5

Table 2
Statistical Reasoning Levels

US 23270	Level of Reasoning Category	US 23271	Level of Reasoning Category
1	1	1	2
2	1	2	2
3	3	3	3
4	2	4	3
5	3	5	3
6	3	6	3
7	2	7	3
8	2	8	2
9	3	9	3
10	2	10	3
11	2	11	3
12	2	12	3
13	2	13	4
14	2	14	4
15	4	15	5

16	2	16	5
17	5		
18	3		
19	3		

Table 3
Statistical Reasoning Levels
Level 4 Unit Standards

	US 23270			US 23268		
	Days	Re-sit Days	Total	Days	Re-sit Days	Total
	8	0	8	6	1	7
	143	97	240	4	76	80
	31	30	61	67	3	70
	39	0	39	84	90	174
	10	5	15	10	1	11
	16	4	20	9	1	10
	21	0	21	26	3	29
	22	5	27	70	2	72
	5	6	11	14	1	15
	5	0	5	14	0	14
	15	0	15	66	30	96
	31	0	31	33	0	33
				23	0	23
Mean	28.8	12.3	41.1	32.8	16.0	48.8
Est Standard Dev	37.6	28.0	64.5	28.5	30.9	48.5
Median	18.5	2.0	20.5	23	1	29

Table 18
Summary Statistics of Completion Times

Level 5 Unit Standards

US 23269			US 23271		
Days	Re-sit Days	Total	Days	Re-sit Days	Total
7	1	8	15	1	16
144	6	150	28	1	29
130	0	130	32	1	33
20	1	21	18	1	19
11	1	12	166	1	167
40	1	41	10	0	10
21	0	21	4	0	4

	26	8	34	108	0	108
	133	0	133	20	0	20
	93	0	93	44	8	52
	71	1	72			
Mean	59.4	1.7	61.1	44.5	1.3	45.8
Est Standard Dev	52.4	2.6	52.6	51.9	2.4	52.0
Median	33.0	1.0	37.5	24.0	1.0	24.5

Table 19
Summary Statistics of Completion Times

Unit Standard 70

1. What are the main objective(s) of the survey?

2. List the sources of data used in this survey and if applicable any surveys that this survey provides data for.

3. How would you classify this type of survey? Choose from one-off, time series or longitudinal. Justify both the type chosen and those not chosen.

4. Explain the difference between a sample and the population from which the sample is selected?

5. What sampling procedure has been used and why was that procedure used?

6. How has the sampling been carried out?

7. Explain the difference between sampling and non-sampling errors.

8. How can a set of observations be weighted?

9. Give an example of each of the four different measuring scales where applicable.

10. Explain the difference between the mean and the median.

11. Explain the difference between the standard deviation and the range.

12. What does the margin of error represent?

13. Explain the meaning of the 95% (99%) confidence interval for the mean and a total.

OR
Explain the meaning of the 95% (99%) confidence interval for a proportion.

14. Explain the meaning of the 95% (99%) confidence interval for the difference between two means (or proportions).

15. Explain how non-responses and missing data have been allowed for by imputation and describe the possible impact of this treatment.

16. Explain how a confidence interval can be used to test an inference.

17. Explain the difference between stratification and clustering and describe the impact of these on reporting the results of your survey

18. Explain how confidence intervals can be used to draw overall conclusions.

19. Explain how statistical measures can be used to draw overall conclusions.

Unit Standard 71
Element 1 Identify, describe and assess legal and ethical constraints on the collection and use of data in a state sector context

1. In your own words give a general overview of the twelve principles stated in the Privacy Act.

2. What are the ethics of using an observational study to collect data from people who are not aware that they are being observed?

3. How are respondents to these data collection(s)] made aware of the purpose of supplying information and/or of the use that will be made of their data?

4. Give your interpretation of how two of the above Privacy principles apply to the collection and use of data in the public sector.

5. Are respondents [to these data collection(s)] made aware of the measures that will be used to ensure their data is secure? If not, suggest how this could have been done.

6. Explain to someone with less statistics knowledge than yourself two security requirements for public sector data collections.

7. Discuss the adequacy of the privacy and security measures applied to the supplied data collection(s) OR, If these are not described discuss what measures do you consider should be applied.

8. How does confidentiality differ from privacy and what is the overall purpose of confidentialising data?

9. Explain to someone who has less statistics knowledge than you, two methods for confidentialising data.

10. How has the data been protected from accidental disclosure of personal information in the published result(s) OR, If it has not been protected, suggest how this data could be protected?

11. Who can access either the original data set or more detailed result? OR, if this has not been identified, suggest who should be allowed access to each level of data (detailed aggregates, original data set, etc.).

12. In your opinion, is the confidentiality protection applied to this data collection(s) adequate? Write this as a briefing note for your manager to consider.

13. What other ethical considerations have been
OR could/ should have been considered with respect to the collection and use of data [in these data collection(s)]?

Element 2 Assess the impact on respondents of legal and ethical issues

14. What measures of response are reported on OR could have been reported on [in these data collection(s)]?

15. Explain to someone with less statistics knowledge than yourself how the quality of the results produced from the data collection(s) supplied could be affected by ethical and legal issues? Discuss at least two potential impacts on respondents.

Element 3 Make and communicate policy recommendations that could help resolve the impact on respondents of privacy, security, confidentiality and ethical issues

16. What changes to the processes related to ethical and legal issues [used in the data collection(s)] could result in reducing the impact of these issues on respondents? Communicate your answer as a recommendation for senior managers to consider.

R

MATHEMATICS TEACHING-RESEARCH JOURNAL ONLINE
VOL 8, N 1-2
Fall and Winter 2015/16

Unit Standard 68

1. Reference the report(s).

2. What are the main objective(s) of the survey?

3. Name a likely non-sampling error resulting from this survey.

4. Identify all the variables used in this survey

5. Classify all the variables identified as continuous, categorical or neither.

6. What statistical procedure(s) have been used and why is it appropriate in this report(s)?

7. Choose a table of counts (frequency tables) and from that table identify and interpret in context three (row, column and total) percentages and/or proportions.

8. Choose a graph or graphs and at least one other type of numerical summary to draw conclusions appropriate to the context of the survey report(s).

9. Choose a statistical measure(s) and/or distribution from the demographic data to draw conclusions appropriate to the context of the survey report(s).

10. Choose a bi-variate analysis and make three interpretations from plot(s), model and/or correlation measure.

11. Give an example of a performance indicator in the context of the survey report.

12. State what a residual is in linear regression.

13. In the context of the survey report, explain what is meant by cyclical variation in a Time Series?

14. In the context of the survey report what could be conclude from the trend in a Time Series?

15. In the context of the survey report, explain what is meant by the random component in a Time Series?

16. How do we interpret a seasonally adjusted time series?

17. Calculate and explain an odds ratio.

18. Calculate and explain an index number.

Unit Standard 69

Element 1: Assess data collections relevant to a policy questions

1. How well do the objectives of the data collection fit the policy question(s)?

2. What is the population that data is required for?

3. If a sampling procedure has been used, describe its main features.

4. What information is being collected?

5. What topics (objectives) do the questions relate to?

6. What aspects of the data collection or analysis affect its use?

7. What is one main result from this study (answer in the form of a possible headline for an article)?

Element 2 Identify and select relevant data collections to make policy recommendations

8. What data collections have you identified that could be used to answer this question?

9. How can these data collections be used to answer the policy question?

Element 3 Describe a statistical information collection that can be used to answer a specific policy question

10. Give the key elements of a data collection that could be used to answer the given policy question?

11. Describe one possible main result (as EITHER a potential press release OR a newspaper article).

12. Describe what sort of graph(s) you could include in your article (in question 11).

Element 4 Evaluate reports based on statistical information to make policy recommendations

13. What constraints are there on the stated conclusions in the supplied articles(s) or report(s)?

14. Do the supplied report(s) or analyses use data that is appropriate to answer the policy question given in element 2? Write your answer so that it could be understood by someone with less statistics knowledge than you.

Element 5 Use reports based on statistical information to make policy recommendations

15. What results or statements made in the supplied report(s) are relevant to the policy question given in element 2?

16. How does the supplied data collection(s) differ from the one you proposed in question 10?

17. How could the data collection(s) have been used to answer this question? OR, How could the data collection be changed so that it could be used to answer this question?

18. Use the result(s) of the analyses described in question 17 to make a policy recommendation or recommendations for your manager to consider.