Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.

THE IMPACT OF CURRICULUM CHANGE ON THE TEACHING AND LEARNING OF TIME SERIES

RACHEL PASSMORE

A thesis submitted in fulfilment of the requirements for the degree Master of Science in Mathematics, The University of Auckland, 2016.

The greatest value of a picture is when it forces us to notice what we never expected to see.

John Tukey

Abstract

The secondary school statistics curriculum in New Zealand has experienced substantial change since 2010. The catalyst for these changes was a desire to improve students' statistical reasoning and to narrow the gap between the statistics taught in secondary school classrooms and the practices and thinking of professional statisticians. Anecdotal evidence suggested the quality of Year 13 student work in time series had improved. This research, therefore, sought to provide a robust analysis of changes in learning outcomes of Year 13 students in time series, in order to determine whether these claims could be supported. Furthermore a literature review produced no evidence of any prior research in the area of student reasoning with time series at this level.

Ethical and time considerations prevented access to samples of student NCEA work on time series completed before and after the curriculum change. Instead exemplars of student work distributed by the New Zealand Qualifications Authority (NZQA), which were freely available on the internet, were utilised. In total 35 exemplars of student work were analysed. In order to obtain a more holistic perspective of the curriculum change, 18 teachers were surveyed and five were interviewed. These secondary data sources provided data about teachers' perceptions of the curriculum change. Since a framework for assessing student learning outcomes for time series did not exist, a framework was developed based on the student data and a synthesis of established frameworks concerned with levels and development of mathematical reasoning, dimensions of statistical reasoning and interpretation of data and data displays.

Analysis of student exemplars against the framework provided strong evidence that after the curriculum change higher levels of reasoning were observed. The shift towards higher levels of reasoning was observed at all levels of achievement – Achieved, Merit and Excellence. The style of student exemplars changed to include a more complete report style response, integrating other research findings that either confirmed or refuted the student's own findings. One of the major facilitators of this change was the availability of free data visualisation software, which liberated teaching and assessment time from a focus on procedures to one of data interpretation and interrogation. The implication of these findings suggest that the framework developed for this study could be used by teachers in order to scaffold their students' reasoning to higher levels and that other NCEA Achievement Standards could be similarly scrutinised in order to evaluate the effect of curriculum change.

Acknowledgements

It has been an absolute privilege to have Associate Professor Maxine Pfannkuch as my supervisor on this project. Whenever I was at a loss to know what the next steps should be, a conversation with Maxine would ensure a suitable path forward was found. She has encouraged me on every step of the way towards completion of this research and I cannot thank her enough for having faith in my ability to complete the task.

I would also like to thank my ever patient husband, Stephen, who has unfailingly supported me throughout many years of part-time post-graduate study and has never once complained. My three children, Sam, Michael and Emma have also had to put up with taking over chores when Mum was trying to finish a chapter.

I am also very grateful to the teachers who completed my survey when I know they probably had a million other things to do. Particular thanks go to the five teachers who allowed me to interview them for an hour and who provided wonderful personal insights of curriculum change from their perspective.

Finally I would like to thank Mike Forster who shares my passion for time series analysis and is always willing to provide a sounding-board about any time series related topic.

Table of contents

Contents

List of figures

List of tables

CHAPTER ONE

Introduction

1.1 Introduction and background

Time series analysis was first introduced into the New Zealand secondary school curriculum in 1996. The topic has been assessed under three different regimens during that time but the most recent change in teaching and assessment is arguably the largest and most significant. Anecdotal evidence from teachers has indicated that the latest time series standard, Achievement Standard 3.8, AS 3.8, (New Zealand Qualifications Authority, 2013) first assessed during 2013, has resulted in a huge improvement in the quality of work and level of reasoning demonstrated by students. The initial motivation for this research was therefore to assess whether this anecdotal evidence could be confirmed or refuted when investigated using robust research methods.

In order to pursue this research question, student work produced under each standard would need to be assessed against some form of independently produced framework in order to provide an unbiased analysis of the level of reasoning found. An extensive search of research to date failed to discover the existence of any such framework. In fact very little evidence of any research in the area of reasoning with time series was found at all. As a result of the absence of any existing framework that could be used for the analysis of student work, a framework was developed as part of this research. This framework has the potential to make a unique and valuable contribution to the body of knowledge derived from statistics education research.

In an attempt to provide a more holistic view of the change in the teaching and assessment of time series in secondary schools, not only were exemplars of student work examined but also a number of teachers were surveyed and a small number of teachers were interviewed. Information collected from surveys and interviews provided background and greater insight into how the change to the new standard had been implemented and received by students and teachers.

1.2 The need for research

The change in the way time series is taught and assessed in New Zealand secondary schools is part of a larger shift in the statistics curriculum. This shift has been prompted not only by technological developments but also by a need to reduce the gap between statistical methods and ideas taught at a school level and those being used by practising statisticians. Another aim of the changes in the statistics curriculum as a whole was to improve the level of statistical reasoning and to promote higher order thinking skills in today's students. In order to assess to what degree these aims have been met requires robust pieces of research conducted under strict research guidelines. Such pieces of research will have obvious interest to the NZQA who have introduced the new statistics standards, statistics educators who have been the catalyst contributing to the change and to teachers who want confirmation that the work involved in implementing these changes is a worthwhile exercise for their students.

A literature review also revealed that very little research has been conducted around the topic of time series analysis, particularly time series analysis taught at the secondary school level. In many countries time series analysis is only taught as a specific topic at undergraduate or post graduate levels. Some journals have featured articles whose focus has been the use of different pieces of software in the teaching of time series analysis or investigations of particularly interesting time series data sets, but no evidence of research on the development of reasoning skills for the analysis of time series was found. This research consequently attempts to fill a void discovered in statistics education research.

1.3 The research question

The main research question was to establish:

1. How, if at all, have the learning outcomes for the new time series Achievement Standard, AS 3.8, changed when compared to the old Achievement Standard, AS 3.1.?

   In order to gain a more holistic insight into the change two further research questions were investigated.

2. What types and levels of reasoning are demonstrated in the two standards and how do they compare?

3. What impact was there on teaching practice with the introduction of the new standard?

In order to answer these three research questions, three sources of data were required. To answer the main research question student exemplars of responses to the two Achievement Standards were examined. To answer the second research question a framework for the level of reasoning demonstrated in time series was developed against which the student exemplars were examined. The third research question was investigated by surveying and interviewing teachers of Year 13 statistics.

1.4 Overview of chapters

This chapter provides an introduction to research submitted in fulfillment of the requirements for the degree of Master of Science at the University of Auckland.

In order to provide an historical background to the changes in the New Zealand statistics curriculum, one of which is the focus of this research, Chapter Two summarises some of the influences on statistics since its introduction to the secondary school curriculum in 1971. Any curriculum change on a national level demands a huge effort from interested and motivated parties. Some of the key characters in the story of statistics education in New Zealand are acknowledged in this chapter as well as some influential statistics educators from overseas who have also played a significant role in shaping the statistics curriculum as it is today.

One major factor in the change in the teaching and assessment of time series was technological. Chapter Three examines the changing role of technology in the mathematics and statistics classroom in general and then in particular how it has affected the teaching and assessment of time series. The development of *iNZight*, (https://www.stat.auckland.ac.nz/~wild/iNZight/), a freely available statistical data analysis package, has played an important role in the success of curriculum change in New Zealand. This chapter examines critical elements required for successful technology integration and investigates how many of these elements were present during the introduction of *iNZight*.

Three different data collection processes were involved in this research. One of them was a collection of student exemplars completed for AS 3.1 and AS 3.8. Chapter Four describes a synthesised framework that was developed in order to ascertain whether the exemplars, completed for two different standards, exhibited any change in the level of student reasoning. The framework draws on a number of other previously established frameworks developed to assess levels of reasoning in assessment tasks, development of mathematical thinking and the development and dimensions of statistical thinking.

Chapter Five describes research methods commonly used in mathematics and statistics education research followed by a detailed account of the actual methods used in this research.

Issues concerning data quality and reliability are also covered in this chapter as are any ethical concerns.

Chapter Six presents the results of this research from the three data sources. In many instances the data from one source supports that collected from the other sources, but a few differences were found and these are highlighted.

Chapter Seven discusses the results presented in Chapter Six and makes some recommendations for the future direction of the teaching and assessment of time series in New Zealand.

CHAPTER TWO

A brief history of the secondary school statistics curriculum in New Zealand

2.1 Introduction

Any current secondary school teacher of statistics in New Zealand will confirm that there have been substantial changes to the school statistics curriculum in the last few years. However, not all of these teachers may be able to explain why these changes have occurred or the motivation behind them. This chapter will explain the rationale behind the reforms by tracking historical changes in the statistics curriculum from the 1950s to the present day. Statistical reform is not unique to secondary schools and has taken place in University courses, not only in New Zealand but also elsewhere in the world. The motivation for change within New Zealand has not only come from those concerned about statistics education but has also been influenced by statisticians in other parts of the world and those working in other disciplines who have recognised the need for improved statistical skills. The statistics curriculum in New Zealand is recognised as innovative and described by Jane Watson, an internationally renowned statistics education researcher as "a leader in the world" (cited in Begg & Pfannkuch, 2004, p12).

2.2 Brief historical overview of New Zealand statistics curriculum, 1950 – 1980

In the 1950s in New Zealand, statistics was not covered in the secondary school curriculum. Students were taught a mathematics-focussed curriculum and did not meet statistics unless they enrolled in an introductory course at University (Forbes, 2014). In the 1960s, Geoff Jowett, the first Professor of Statistics at Otago University mentioned proposals to include some statistics content at a senior secondary school level. Jowett was well-known for his practical demonstrations of statistical concepts and indeed built a number of devices. Such an illustrative practical approach to learning statistics was unusual, if not, unique at that time. Jowett and another applied statistician, Stan Roberts became advocates for change in the secondary school curriculum and in 1971, a secondary school subject called Applied Mathematics was established. This offered students a choice of three options – mechanics, statistics and computing – of which they had to choose two. Roberts (1999, p. 2, cited in Forbes, 2014) described this as a "turning-point for statistics in secondary schools".

By 1980 and fuelled by further statistics education activists such as David Vere-Jones, the two secondary school topics of *Mathematics* and *Applied Mathematics* were replaced with

Mathematics with Calculus and *Mathematics with Statistics*, but again these courses were designed only for senior secondary school students. For the first time an internally assessed practical investigation or project was introduced, but not universally welcomed by teachers of the day and it was eventually dropped from the curriculum. However, in their defence, many of these teachers had little or no professional training in statistics and were understandably nervous of this change. Further details about key players during this time are available in Forbes (2014).

2.3 1980 – 1992, New Zealand

In the 1980s, the New Zealand Statistical Association, (NZSA), and in particular its Education Committee, became actively involved in school statistics education. By now significant advances in technology were being made and this committee sought to assess how these advances might impact the school statistics curriculum. NZSA members were active lobbyists and advisors throughout the development of the 1992 mathematics curriculum that included a statistics strand for all levels of schooling for the first time.

2.4 1992 – present day, New Zealand

The 1992 curriculum (Ministry of Education, 1992) contained six strands of which statistics was one. As only one of six strands, statistics teaching took up about six of the forty teaching weeks per academic year. The 1992 curriculum introduced achievement objectives (information about student learning experiences) and in the statistics strand these included – statistical investigations, interpretation of statistical reports and exploration of probability.

In 1995 an addendum to the final curriculum was published (Camden, 2003) which added three further statistical topics to the curriculum – time series, experimental design and bivariate data analysis. Time series was introduced before the other topics and was first assessed in 1996.

In 2001, after the 1992 curriculum had been in place for a decade, the Ministry of Education instigated a review. A conference of all interested parties, including professional statisticians, university lecturers in statistics, experienced teachers and teacher educators, was held. This group suggested an extensive literature review be undertaken which was led by Maxine Pfannkuch and Andy Begg (2004). This review took the form of a number of recommendations about a variety of aspects of the school statistics curriculum and was made after consultation with international experts in statistical education. Some ideas and terms used remain important challenges today, for example, the role of technology, the role of context in statistical thinking, statistical literacy and the importance of communication.

In 2007, a new mathematics curriculum with a name change to the mathematics and statistics curriculum was finalised, (Ministry of Education, 2007). The number of strands was reduced to three, Number and Algebra, Measurement and Geometry and lastly Statistics as a separate strand in its own right. The recommendation was that Statistics should gradually increase in importance as a student moved through the year levels at school. By senior secondary school statistics should be taught for about 13 weeks of the year as opposed to only six weeks in the 1992 curriculum. This doubling of the recommended teaching time meant that entire course schemes had to be reviewed and rewritten.

The statistics curriculum was now firmly based in data-handling rather than mathematics. For the first time new computer-based techniques such as bootstrapping and randomisation methods were proposed and new standards such as evaluation of statistical reports introduced. Given the changes to the statistics curriculum and other curricula, a phased introduction was implemented, with changes to the National Certificate of Educational Achievement (NCEA) Level 1 being made in 2011, to Level 2 in 2012 and to Level 3 in 2013. The new curriculum offered a large number of statistics Achievement Standards, including some using the new techniques and requiring the use of technology, from which teachers could tailor their own statistics courses. Furthermore by senior secondary school (Year 12 and Year 13) it was now possible for students to do courses solely in statistics rather than a mixture of mathematics and statistics topics. Current University of Auckland staff members, Chris Wild and Maxine Pfannkuch were leaders in the development of teaching resources for the new Achievement Standards, resulting from large-scale research projects and the development of a free data visualisation tool called *iNZight*.

Also introduced in the 2007 curriculum was the notion of key competencies. These are listed as:-

1. Thinking
2. Using language, symbols and texts
3. Managing self
4. Relating to others
5. Participating and contributing

These key competencies were viewed as essential characteristics of life-long learners and contained many qualities that were central to the new way of teaching statistics. For example, developing intellectual curiosity, using technology to communicate with others and interpretation of information in a range of contexts; all important skills for statisticians, which can be viewed as demonstrations of these key competencies.

Since 2013 was the first year of implementation for the Level 3 standards it is too soon to judge yet what the eventual uptake will be of the Achievement Standards that incorporate new techniques such as bootstrapping and randomisation methods or those that require data-visualisation software. These are ground breaking changes in secondary school curricula; no other country has such a forward looking curriculum. Success will almost inevitably depend on the priority individual schools give to the change as well as the motivation of individual teachers and Heads of Department to try something new and unfamiliar. Without a substantial investment in professional development of teachers and of teaching and assessment resources the likelihood of widespread implementation may not be high (Patterson and Czaikowski, 1979).

2.5 International influences

Many international influences have affected the secondary school statistics curriculum in New Zealand and indeed in other countries too. This section will identify the influences of some but definitely not all of those whose contribution has been significant. The earliest influence is probably that of Tukey (1977) whose work on exploratory data analysis had extensive impact. Exploratory data analysis (EDA) was used by Tukey to describe data in a simple comprehensible manner in order to reveal stories behind the data without making any underlying assumptions about the data. Paramount in his descriptions was the use of a variety of data display techniques – a precursor to the data visualisation techniques that technology permits today.

Also in the 1970s another influential statistician was at work. In 1979 Brad Efron published a paper entitled *Bootstrap Methods: Another Look at the Jackknife.* Although the paper was published 35 years ago and despite acknowledgement of its importance from Efron's peers, it is only now that bootstrapping methods have infiltrated the New Zealand secondary school curriculum and indeed was only recently introduced into some tertiary curricula. Today's secondary school teachers may question why, if bootstrapping was such a good idea, has it taken 35 years to reach school curricula? The answer is complex but among the relevant factors are that at the time of the paper's publication theoretical statisticians were beginning to feel somewhat threatened and therefore defensive of their school of thought. Technology was new and changing quickly, making some people question methods that had simply not been possible before the advent of computing power. Even Efron (2000, p. 1295) later admitted that "it has taken me a long time to get over this feeling that there is something magically powerful about formulas…. and to start trusting in the efficiency of computer-based methods".

During the 1990s calls for reform in statistical education were coming from a number of different quarters. Katherine Wallman (1993, p. 1) in her address to the 1992 annual meeting of the American Statistical Association, noted "the importance of strengthening understanding of statistics and statistical thinking among all sectors of our population". Gal and Garfield (1997, p. 1) mention that "Statistics has gained recognition as an important component of the precollege mathematics and science curriculum". The increased importance of statistics was also being recognised in other disciplines. Sokal and Rohlf (2012, p. 5) conducted a survey of papers that appeared in *The American Naturalist* since 1890. They discovered that few papers contained any statistics up to 1940, but since then there had been a dramatic increase. They note that "By 2000, 97% of the papers published…..employed some statistical tests or were more heavily statistical or mathematical. This trend can be observed in most biological journals". Biology was not the only discipline calling for improved statistical outcomes for their students. The combination of the improved profile of statistics and the call from a variety of disciplines for students to have sound statistical knowledge led to many researchers investigating new methods of teaching and assessing statistics at tertiary and secondary level. Gal and Garfield (1997) challenged traditional methods of teaching statistics within mathematics and noted five key differences between mathematics and statistics using this to create a platform for teaching and assessment reform. In particular they suggested that new approaches must integrate student skills and knowledge with abilities to interpret and communicate responses to contextual problems. They recommended the use of projects, extensive use of technology and portfolios as possible forms of assessment.

In 2005 another influential report was produced, Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (Franklin, Kader, Mewborn, Moreno, Peck & Scheaffer, 2005). The goal of this report was unashamedly to improve levels of statistical literacy.

> Our lives are governed by numbers. Every high-school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family and to be prepared for a healthy, happy, and productive life. (GAISE, p. 1)

The problem was that as statistics used to be taught, firmly rooted in assumptions about theoretical distributions; it was inaccessible to a large cohort of students. This was no longer acceptable given the call from a range of disciplines that needed students who could think statistically and handle quantities of data effectively.

In 2007, George Cobb published *The Introductory Statistics Course: a Ptolemaic Curriculum*. In this paper Cobb (2007, p. 1) notes that

> Our generation is the first ever to have the computing power to rely on the most direct approach… Just as computers have freed us to analyse real data sets, with more emphasis on interpretation and less on how to crunch numbers, computers have freed us to simplify our curriculum.

He reiterates comments made fifteen years earlier by David Moore (1992, cited in Cobb, 2007, p. 1): "we can and should automate calculations; we can and should automate graphics". Cobb goes on to make the seemingly radical statement that "now we have computers, I think a large chunk of what is in introductory statistics courses should go" (Cobb, 2007, p. 4). Cobb recommends a new statistics curriculum that is based not on distribution based methods but on the logic of inference. The new method is summarised by what he terms the three Rs: randomise, repeat, reject. Above all a new statistics curriculum should be easily accessible to the majority of students. Before the advent of computing power there were no choices in how statistics was taught, but with it the excuse for teaching statistics centred around the Normal distribution has lost any validity it may once have had.

2.6 Impact of the reform on the teaching of time series

In New Zealand the teaching of time series analysis is taught in Year 13, the final year of school. It was first taught as part of the Bursary Statistics course in 1996 and was last assessed in this form in 2003. The Bursary course was replaced by NCEA Level 3, and the topic of time series became Achievement Standard 3.1, first assessed in 2004. This was a popular internally assessed standard with which large numbers of students achieved assessment success. However, although the data requiring analysis was offered in a contextual setting, students were not required to engage particularly with the context but were more focussed on producing sets of calculations based on the original data. Achievement Standard 3.8 replaced Achievement Standard 3.1 in 2013 and was the first attempt to reflect some of the recommendations made by statistical education reformers in the topic of time series analysis. Precise details of these changes will be examined in a later chapter, but suffice to say now that the new standard shifted the focus away from repetitive calculations and towards interpretation of data visualisations produced using technology. Exploratory data analysis, as defined by Tukey, involves "detective work….the analyst of data needs both tools and understanding" (Tukey, 1977, p. 1). These fundamental principles of exploratory data analysis are reflected in this new Achievement Standard. The call, by Gal and Garfield (1997), for students to handle, interpret and communicate about data sets from meaningful contexts, as stated in the curriculum documentation, is also reflected in this new standard. Moore's (1992, cited in Cobb, 2007) insistence on the automation of calculations and production of graphics is also present with the requirement that this standard be assessed using current technology such as the software packages *iNZight* and *GenStat* (https://www.vsni.co.uk/downloads/genstat/). Finally, as predicted by Cobb (2007), computing power has enabled the time series analysis curriculum to be freed from tedious calculations and shift towards deeper interpretation and insight of the data.

CHAPTER THREE

The changing role of technology in the mathematics and statistics classroom

3.1 Introduction

When I first returned to a secondary school mathematics classroom in 2000 after a break of 23 years, my first impression was surprise that the classroom environment was remarkably similar to the one I had experienced as a student. Text-books were in use with familiar content layouts followed by worked examples and lots of exercises. All students had scientific calculators and, like I had been, were allowed to use these in assessments. More whiteboards rather than black-boards were in evidence but otherwise the classroom environment looked remarkably similar. So why did this lack of change come as a surprise? I was aware of considerable media hype about the dramatic technological changes happening in society and had presumed that this would be reflected in the secondary school mathematics and statistics classroom – but there was no computer of any description in sight. The school did have computer laboratories that I, as a teacher of mathematics, could book, but it was not encouraged and I found my colleagues rarely utilised computers at all in their teaching.

However, today's mathematics and statistics classroom is a different place. Chance, Ben-Zvi, Garfield and Medina, (2007, p. 1) suggest

> Today's statistics classes may be taught in a classroom with a computer projected on a screen, or may take place in a laboratory with students working at their own computers. Students commonly own a calculator more powerful than the computers of 20 years ago. Others may use a portable computer (laptop) at school, home and on the move.

Several hundred schools in New Zealand, including primary, intermediate and secondary schools are operating what has become known as a Bring Your Own Device, (BYOD), policy which encourages students to bring their laptop, tablet, smart phone or other technological device to school (Irwin & Jones, 2014). The BYOD policy has been promoted on the Te Kete Ipurangi (TKI) website which suggests that a

> learner-centred curriculum that includes digital devices supports greater flexibility in learning pathways, empowering students to learn in a more personalised way with increased control over their own learning ("Learning with 1:1 digital devices", 2014).

For some schools, BYOD has been an optional policy but increasingly schools are making bringing a device compulsory. Issues of compatibility have led to some schools specifying the kind of device students must have, but others have allowed any device reasoning that the benefits of students bringing a device to school outweigh the problems involved in the resolution of compatibility issues. BYOD policies are not the exclusive right of the more wealthy schools either – this policy has been introduced in schools from across the economic spectrum. BYOD has increased in popularity through increased personal ownership of devices, which has meant that schools do not necessarily need to shoulder the financial burden of provision.

In addition to students bringing their own devices to school the classroom has also begun to incorporate other items of technology. For example, data show projectors have largely replaced the old overhead projectors. Smart boards have also been introduced but these are not as widely used as other items of technology, largely because they still represent quite a major financial investment.

This chapter will consider the benefits and pitfalls of the introduction of technology and examine in particular the part technology plays in the statistics classroom. Some software developments that have impacted on the statistics classroom and curriculum will be considered as well as how these developments have affected the teaching and assessment of time series in secondary schools. Finally an analysis of research conducted in the last fifteen years on time series analysis will be presented.

3.2 Integration of technology

Technology enthusiasts, particularly in the early days of technological reforms, were quick to voice the benefits of technology in the teaching of mathematics. However, some researchers, (e.g., Lynch, 2006; Berger, 1998), questioned this common assumption and noted that government policies, which focussed on equipping school classrooms with technological devices alone, were unlikely to promote any beneficial improvements in the learning of mathematics. Mathematics educators quickly realised that technology was not going to be the panacea that was initially anticipated and consequently a considerable body of research emerged to investigate how technology might be successfully incorporated into the mathematics and statistics classroom and curriculum. A brief review of some of this extensive area of research is now presented.

3.2.1 Advantages and disadvantages of technology integration

Advocates of technology suggested that its introduction had the potential to enhance learning and that it could enable change in pedagogical style and delivery of the curriculum. Smith

(1998) discusses the Kolb learning cycle which categorises learning in the following four stages:

1. Concrete experience
2. Active experimentation
3. Reflection and observation
4. Abstract conceptualisation

Smith showed how technology could be used to support each of these learning stages in the teaching of calculus. Experience of data plots gave students concrete experience, ease of graphical representations supported active experimentation, and the saving of time through automation of those graphical representations encouraged students to reflect on their meanings. Not all students reached the fourth stage of abstract conceptualisation.

Researchers, (e.g. Pea, 1985), suggested that technology could be used as a *cognitive amplifier and reorganiser*. They suggested that technology could be used to extend existing curricula by exploiting the new learning opportunities offered by technology. In turn they recognised that this may also necessitate some reorganisation of the curriculum and re-allocation of priorities.

One of the most common learning benefits of technology cited by several researchers, (e.g. Thomas & Chinnappan, 2000) is the ability to link different representations. The ability to switch between graphical, tabular, algebraic and conceptual representations can help to promote mathematical enquiry. However, Thomas and Chinnappan caution that effective student use of technology relies heavily on the skill level of the teacher. Another advantage of technology has been to reduce student cognitive strain (Hillel, 1993). Tall (2000) also mentions this aspect when he talks about using technology so that a student can focus on one particular aspect of theory at a time while other aspects are handled by the software. As each aspect is mastered another can be illuminated again using the software to assist in this process. While some recognise this as an advantage of technology others remain unconvinced of this so-called '*black-box*' approach where students may have no idea of the underlying tasks being performed. Participants in research by Olson and Keynes (2001) suggested that many mathematicians still maintain that knowledge of such underlying processes is essential if certain concepts are to be grasped fully.

A study conducted by Martin (2000) into the benefits of graphical software in the teaching of calculus again did not produce the enhanced learning outcomes that were anticipated. These typically less favourable outcomes concerning the introduction of technology encouraged further research into what emerged to be a complex issue. In particular a considerable amount of research was conducted that considered the interaction of teacher and student with the technology and the importance of recognising what features of these interactions might result in positive or negative benefits from the integration of technology.

Goos, Galbraith, Renshaw and Geiger (2000) considered the interactions of students and teachers with technology separately through the following four categories:-

1. technology as a master
2. technology as a servant
3. technology as a partner
4. technology as extension of self

For example, a student who uses technology as a servant is one who will turn to technology to replace conducting time-consuming procedures by hand. Whereas a student who views technology as a master, may use technology for the same task but have a blind acceptance that whatever the technology produces must be correct. However, the researchers suggest that it is only when students begin to view technology as their partner, do the true learning benefits emerge. As a partner, technology can be used creatively to explore mathematical concepts and data sets. The final category, technology as an extension of self, was seldom observed by the researchers either as a teacher or student interaction with technology. This type of interaction is only attained if technology is integrated into all facets of the teaching or learning of mathematics, that is, pedagogical practice, learning opportunities, assessment and curriculum. It is this level of interaction that is seen as crucial to the successful integration of technology.

Interactions with technology are closely linked with personal attitudes towards technology (Stacey & Pierce, 2001). Students in one study appeared resistant to the use of technology, showing a preference for by-hand techniques and demonstrating a lack of trust in the technology (Stacey & Kendal, 2001). Such reactions frequently stem from the attitudes of the students' teachers. A negative attitude by either the teacher or the student towards technology is unlikely to result in positive learning outcomes.

Such research illustrates that the successful integration of technology into the teaching and learning of mathematics and statistics is not straightforward. Table 1 below from an unpublished doctoral thesis, (Oates, 2009), identifies and describes the factors critical to the successful integration of technology.

Table 1 Critical elements in technology integration

Image removed from final copy due to copyright. Please see citation below.

(G. Oates, Integrated technology in the undergraduate mathematics curriculum: A case study of computer algebra systems, 2009. Reprinted with permission.)

Although these factors were established through a case study concerned with the integration of computer algebra systems, they will also be relevant to the integration of other types of technology. Factors described in Table 1 reinforce the research findings outlined earlier, for example, congruency of access to technology. The BYOD policy prevalent in many schools has sought to provide this and helped to enable technology to be available at school and outside school. Congruency in the use of technology in teaching, learning and assessment, also identified as a critical factor by Oates was seen as important by Smith (1998) and Pea (1985) if the true potential of technology is to be realised. Negative staff attitudes, perhaps because of unfamiliarity with technology, or lack of professional development are noted under staff factors, but were also examined by researchers Stacey and Kendal, (2001) and Stacey and Pierce, (2001), and found to be critical in shaping the attitudes of their students to technology.

3.3 Technology in the statistics classroom

In the statistics classroom, over 35 years ago, Tukey (1977, p. 1) was encouraging statisticians and those teaching statistics to do more with their data. He suggested that they should aspire to be "….a data detective (who searches) for interesting and unexpected results".

Tukey experimented with different types of data displays with the aim of exposing the stories held within the data. Undoubtedly, the ease of data visualisation afforded by new technology has made the role of the data detective much easier and open to a wider range of students than before.

Later, Moore (1997, p. 128), as mentioned in Chapter Two suggests:

> Automate computations and graphics: Automating computation is controversial among mathematicians. It is much less controversial among statisticians, as it reflects the practice of our discipline.

Such calls for automation were echoed by Gal and Garfield (1997) when they defined eight goals for the teaching of statistics. Gal and Garfield suggested that the development of interpretive skills would be important for tomorrow’s students, thus one of their six goals was to develop interpretive skills and statistical literacy. Students, they suggested, will need to practise interpretive skills if in their adult life they are to have the ability to pose critical and reflective questions from data published in the media. The pre-technology statistics curriculum required students to be able to construct, by hand, different types of data displays. The teaching of these construction skills took up a large proportion of teaching time leaving little, if any, time for the development of interpretative skills .The posing of “what if”- type questions was extremely limited, and in many classrooms probably not covered at all. Technology can help to reduce the students’ cognitive load, replacing and/or automating complex procedures, which then allows students to focus on higher-level understanding (Moore, 1997).

However, as detailed in the previous section the simple presence of technological options does not guarantee positive learning outcomes. Chance et al. (2007, p. 8-13) suggested the following ways in which technology could be used to enhance learning in the statistics classroom.

1. Automation of calculations
2. To increase emphasis on data exploration
3. Visualisation of abstract concepts
4. Simulations as a pedagogical tool
5. Investigation of real life problems
6. Provision of tools for collaboration and student involvement

If such enhanced learning opportunities were to be experienced by students then software to meet these needs had to emerge. Traditional packages used by practising statisticians, like *SPSS*, *SAS* and *Minitab*, were not going to fit the brief required by those teaching and learning statistics in the 21st century. Software aimed at the statistical education market needed to be developed and the first two products to fill this void were *Fathom* (http://www.keycurriculum.com/resources/fathom-resources/getting-started-with-fathom) and *Tinkerplots* (http://www.keycurriculum.com/products/tinkerplots). Erickson (2002) describes *Fathom* as a dynamic computer-learning environment for teaching data analysis and statistics based on dragging, visualisation, simulation, and networked collaboration. *Tinkerplots* was aimed at a slightly younger market; primary and intermediate aged students, but used a similarly user-friendly approach. Both *Tinkerplots* and *Fathom* have been widely used in a number of different countries with many positive benefits observed (Ben-Zvi, 2006; Biehler, Ben-Zvi, Bakkar & Makar, 2013; McClain & Cobb, 2001; Pfannkuch, 2008; Pfannkuch & Ben-Zvi, 2011; Konold & Miller, 2005; Finzer, 2006).

Although *Fathom* and *Tinkerplots* were used overseas their implementation in New Zealand schools was not extensive. For many schools the cost was prohibitive. If New Zealand students were to keep up with their counterparts overseas who were benefiting from a new software environment, another solution needed to be found. Two possible solutions emerged, one called *Genstat*, which was originally developed by the Rothamstead Experimental Station in the UK but hosted by the University of Otago in New Zealand and one from the University of Auckland called *iNZight*, (https://www.stat.auckland.ac.nz/~wild/iNZight/) developed by Chris Wild. Both packages were and still are free to download. These packages seem to have had far-reaching effects on the teaching and learning of statistics in New Zealand. In particular positive benefits have been observed anecdotally with the bootstrapping and randomisation modules. In the next section, the effect that one of these packages, *iNZight*, has had on the teaching of time series analysis will be examined. One commentator, Dr Nicola Petty (2013) is extremely enthusiastic about the merits of *iNZight* and even goes so far as to say “For a high school teacher in New Zealand, there is nothing to beat it”.

3.4 Technology in the teaching of time series analysis

The inclusion of time series analysis in the New Zealand statistics curriculum dates back to 1996 when it formed a small part of the Bursary examination. After the introduction of the National Certificate of Educational Achievement (NCEA), the topic of time series became an

Achievement Standard in its own right (AS 3.1). This standard was eventually replaced in 2013 by a new Achievement Standard also exclusively concerned with the analysis of time series (AS 3.8). Not surprisingly technology use over this period of time has changed to reflect new standard requirements but also as a consequence of the increased availability of software and hardware for time series analysis.

The teaching of the time series component of the Bursary examination was heavily based in numerical calculations and the most common item of technology in use at that time was the calculator. With the introduction of Achievement Standard 3.1, for the first time the standard explicitly stated that "*the use of appropriate technology is expected*". The technology most commonly in use then was an *EXCEL* spreadsheet which was used to automate calculation of smoothed data, individual and average seasonal effects, to fit a linear trend and to produce predictions and a graph. Although technology use was *expected* for the first time it was still not used universally. Some students had no experience of *EXCEL* and anecdotal evidence suggests that teachers of statistics were reluctant to use valuable teaching time to familiarise students with *EXCEL*. So it was still possible to complete this Achievement Standard using only a calculator. Student exemplars for AS 3.1, posted on the NZQA website and a focus of this research, demonstrate outputs from *EXCEL* but also tables of calculations that may have been carried out by hand or using a calculator.

With the introduction of the latest Achievement Standard, (AS 3.8), teachers of statistics were left in no doubt that technology was now expected as national exemplars (NZQA, 2013) produced by the NZQA demonstrated how the new free software package, *iNZight*, could be used to analyse time series. The new standard, through the exemplars, not only illustrated the use of software such as *iNZight* in order to examine the decomposition of a time series but it also required students to research the context of their time series and recommended that such research might be conducted using the internet or a library. So for the first time in the teaching of time series technology was being used for tasks other than procedural calculations. This automation of many of the calculations permitted more time for interpretation of a time series within its context and the new data visualisations permitted students to see clearly for the first time the components of a time series.

As previously discussed, for this new Achievement Standard to produce positive learning outcomes the availability of *iNZight* and other free software packages are not in themselves sufficient to effect changes in the teaching and learning of time series. A number of critical factors, described in Table 1 require consideration before a successful widespread introduction of any new technology. I will briefly assess how, if at all, these critical factors were addressed prior to the introduction of the new time series Achievement Standard, AS 3.8.

Organisational Factors. *iNZight* was freely available on all devices. The NZSA education committee (Pfannkuch, 2011) supported the new technology and workshops were held in advance of the introduction of the standard. Support was also available via on-line forums. *iNZight* was promoted not just for use on the new time series standard but also on many of the other new standards. However, this standard was introduced on the back of changes in each of the two previous years so could be criticised for allowing insufficient time for implementation.

Mathematical factors. Some of the new mathematical content was unfamiliar to teachers. For example, the Holt-Winters smoothing technique used by *iNZight* was new to most teachers. Without knowledge of this technique, some teachers might reject the software as a *black-box*. Teachers were provided with a guide to the Holt-Winters smoothing technique posted on the *Census@School* website (Passmore, 2012a), a central depository for many statistical resources.

Assessment. Although exemplars promoted the use of *iNZight* in assessments the precise time management of the assessment was left to schools. A clarification of the standard suggested that multiple sessions might be required but whether these were supervised or not was unclear. Differences in assessment management could lead to a lack of confidence in the standard. The use of *iNZight* in on-line exemplars of student work helped to promote congruency between teaching, learning and assessment.

Access. Allegations of equity problems in access to technology have largely dissipated with almost all secondary schools providing some access. Many schools, as previously mentioned, now operate a BYOD policy on which students could access *iNZight* and the internet for research.

Staff factors. This is probably the most critical factor of all. If teachers do not engage with the new technology and do not make the effort to integrate the technology into their teaching, their learning activities and their assessment opportunities, then the chance of successful learning outcomes becomes unlikely. In preparation for the introduction of this standard, professional development workshops were held throughout the country. Exemplars were available on-line, and websites such as *Census@School* (http://new.censusatschool.org.nz/) contained numerous resources and other support material for teachers.

When examined against these critical factors the introduction of *iNZight* can be viewed as a success story, the evidence of which can be claimed through its widespread usage in secondary schools throughout the country (Forbes, Chapman, Harraway, Stirling, Wild, 2014). Organisationally, there was a lot of support for new users of the software from workshops, professional development events, podcasts and on-line forums. The software was available on all devices although teething problems for Mac users were more common than on other devices. In terms of mathematical factors, time series as a topic was not widely studied by teachers since it was not part of the school curriculum in their day or a commonly taken paper at undergraduate or postgraduate level. Yet teachers had felt comfortable teaching the old standard as it involved many procedural calculations that they did understand. The new technique of smoothing data used in *iNZight* was unknown to most teachers which may have made them feel uncomfortable about using *iNZight* because of their lack of ability to explain exactly what the software was doing. Professional development events in the year leading up to the implementation of the new standard may have gone some way to allaying these fears. Teachers rely heavily on exemplars posted on-line by NZQA, so it was important for the success of *iNZight* for outputs from *iNZight* to appear in these exemplars. Even before the time series Achievement Standard was finalised, exemplars had been produced to show teachers how the new software could be used to teach the new standard (Passmore, 2012b). The belief was that once teachers saw how to incorporate *iNZight* into the assessment process they may be more inclined to try the new software.

Raiti (2007, p. 6) suggests that there is probably one overarching critical factor crucial to the successful implementation of technology and that is the culture of the school environment.

> Making sure that there are sufficient computers in your school won't necessarily encourage your teachers to use new technology. It's your school's learning culture which is more likely to determine whether desired improvements to teaching and learning can be effected through the integration of ICT.

Without this change in culture, Raiti suggests that ICT can become simply an add-on to existing structures and pedagogies. The traditional model of a teacher as an imparter of knowledge needs to change with the implementation of new technology. Technology allows for a more interactive mode of teaching but only if teachers can adapt their existing pedagogical style.

Chapter Four will examine in greater detail the changes that have occurred in New Zealand in the teaching and assessment of time series. I will seek to assess whether the shift in emphasis permitted by the development of new software has resulted in positive learning outcomes and a deeper level of understanding of time series by students and teachers of statistics.

3.5 Research on the teaching of time series

Much of the research conducted on time series examines the use of technology, provides details on possible learning opportunities using that technology or examines interesting data sets. Other time series research is theoretical but few researchers have reflected on the development of reasoning skills required for in depth analysis of time series. Table 2 displays the results of internet searches using time series in the title. Only research conducted in the last 15 years was considered. The categories used in Table 2 are defined as follows:

Theoretical. Papers in this category focussed on particular statistical techniques, tests and refinements of tests or compared problems from different theoretical perspectives.

Use of software tool. Papers in this category described uses of new and existing software. Either particular applications of the software were described or specific examples of how the software could be used to examine data sets were given.

Reasoning skills. Research in this category considered the development of skills required for a variety of statistical topics taught at levels from primary school through to undergraduate levels.

Teaching activity. Papers in this category described in detail teaching activities that the person had used with their own class. An evaluation of the activity in terms of the learning outcomes achieved for their students would usually be included.

Data sets. Papers in this category described interesting data sets from a variety of contexts using a variety of statistical methods for analysis.

Table 2 Analysis of internet searches for research on time series.

Journal	No. hits	Theoretical	Use of software tool	Reasoning skills	Teaching activity	Data sets
Teaching Statistics	12	3	3	0	1	5
Statistics Education Research Journal	4	1	2	0	1	0
Journal of Statistics Education	88	21	19	0	11	37
MathEduc database	33	29	2	0	2	0
International Association of Statistics Education	6	0	3	0	1	2

As Table 2 illustrates no research on the development of reasoning skills for the analysis of time series was found during the Internet search of over 2000 articles on statistics education. This is partly explained by the fact that time series is not part of other countries' school curricula but it is part of most undergraduate statistics curricula. This thesis hopes to contribute to this neglected area of research.

3.6 Summary

This chapter has demonstrated numerous factors that require careful consideration before the implementation of any technological or curriculum change. The transition from the view that all technology must be beneficial, to technology is not always beneficial has been a long journey. Today, as a result of considerable research, critical factors, which impact on the successful implementation of change, have largely been identified; the challenge now is to accommodate them successfully into a programme of reform.

However, even if issues such as availability of hardware and software options, timing, and professional development are resolved the success or failure of any form of curriculum change whether technological, content, pedagogical or otherwise will ultimately be dependent on the teacher in the classroom for its implementation. Handal and Herrington (2003, p. 59) suggested that

> Curriculum change may only occur through sufferance as many teachers are suspicious of reform in mathematics education given its equivocal success over the past decades.

A considerable body of research has been conducted investigating the influence of teachers' beliefs on curriculum change. Many researchers in the 1990s highlighted how teachers' beliefs can result in teachers becoming enablers of change or innovation but alternatively their beliefs can result in the creation of barriers and obstacles for change (Haynes, 1996). As a result protagonists for curriculum change or technological innovation now need to be aware that any change will be problematic for teachers and even intimidating (Martin, 1993). Teachers will not automatically respond to curriculum or technological change with enthusiasm but must be convinced that the change in question is beneficial to the learning outcomes of their students and is in some way superior to their current teaching practice.

Table 2 shows that very little research has been conducted on the benefits of using technology to improve higher order thinking skills in the teaching of time series. Unless more research is conducted which provides evidence of the benefits of such a change in the taught curriculum, widespread pedagogical, curriculum and assessment reform is unlikely to eventuate or be successful long term.

CHAPTER FOUR

A framework for types and levels of statistical reasoning with time series

4.1 Introduction

In this chapter the development of statistical reasoning will be examined through a number of different theoretical lenses. The frameworks discussed have slightly different foci but elements from each of them will be utilised to establish a synthesised framework for the types and levels of statistical reasoning with time series. The synthesised framework will be used subsequently to determine the level of statistical reasoning achieved in exemplars of student work completed for the now expired Achievement Standard, AS 3.1, and the current Achievement Standard, AS 3.8. Although a framework for the development of reasoning in time series has not been developed, synthesised frameworks for other areas have been developed. In particular, Mooney, Langrall and Hertel (2014), developed a synthesised framework for probabilistic thinking. It is hoped that the time series framework, developed as part of this research, will not only enable the research question to be answered but may also help to inform instruction techniques and assessment writing. The framework may also be used to examine how current and future curricula reflect the different levels of reasoning described.

4.2 Theoretical perspectives

There are several theoretical lenses that offer a variety of ways to examine the development of mathematical and statistical reasoning. I have categorised these lenses into three distinct groups:

1. Frameworks developed for analysis of levels of reasoning in assessment tasks
2. Frameworks characterising the development of mathematical thinking
3. Frameworks characterising the development and dimensions of statistical thinking

A full review of frameworks in these categories is outside the scope of this thesis. Each framework selected is briefly described and its key components summarised in Table 3. An examination across the frameworks in all categories provided a synthesis of frameworks which is summarised in Table 4.

4.2.1 Two frameworks developed for analysis of levels of reasoning in assessment tasks

The frameworks selected in this section to inform the development of the time series framework are De Lange’s assessment pyramid and the Structure of the Observed Learning Outcome (SOLO) taxonomy. De Lange’s assessment pyramid was developed within the Realistic Mathematics Education (RME) school of thought. The RME was selected because it focuses on the levels of reasoning which can be instigated through the design of an assessment task and it considers the role of context, crucial in any statistical reasoning, in solving problems. The SOLO taxonomy was selected because it focuses on the levels of student responses to a task. SOLO is also widely used in statistics education research (Watson, 1994) and is also used in New Zealand’s NCEA assessments to describe the different levels of thinking in the mathematics and statistics Achievement Standards.

Realistic Mathematics Education

RME has been developed over the last 35 years by researchers at the Freudenthal Institute in The Netherlands (Heuvel-Panhuizen, 1996). A strong focus of the curriculum developed by this Institute is the need to provide realistic contexts for students to emphasise the usefulness of mathematics in the real world. This involves developing students’ ability to mathematise real world situations. Students following a realistic mathematics course are also expected to develop a critical stance and evaluate when mathematics has been used to intimidate or even deliberately mislead. In order to assess how well an activity or assessment task reflected the components of an RME style, De Lange developed a pyramidal framework illustrated in Figure 1. (Verhage & De Lange, 1997)

Image removed from final copy due to copyright. Please see citation below.

Figure 1 De Lange's assessment pyramid.

Reprinted from Verhage, H. & De Lange, J. (1997, April). Mathematics education and assessment. *Pythagoras*, p 14-20. Copyright 1997 by Association for mathematics education of South Africa. Reprinted with permission.

There are several components to De Lange's pyramid, namely

1. *Levels of reasoning or thinking.* Lower level involving reproduction, a middle level requiring connections to be made and a higher level requiring analytical skills.
2. *Context.* Context free, camouflage and authentic.
3. *Complexity.* Ranges from simple to complex.
4. *Domains.* Algebra, geometry, number and statistics & probability.

A task which requires a lower level of reasoning would be one concerned with basic facts, procedures and recall of knowledge. Questions in such tasks would be similar to ones previously practised and would require little if any new level of reasoning or thinking to achieve. A task which requires a middle level of reasoning would require the student to integrate information from perhaps a number of sources in order to develop a strategy for solution. At the middle level of reasoning a student will need to apply previously learned skills in a new way or to a new and perhaps unfamiliar context. At the highest level of reasoning students would be expected to demonstrate many of the following characteristics – critical attitude, interpretation, reflection, creativity, generalisation and mathematisation. The process of mathematisation is considered to have two components – horizontal and vertical. Horizontal mathematisation requires students to convert a realistic context into a mathematical problem that can be solved mathematically. The processes involved in the mathematical solution are described as the vertical component of mathematisation. A student displaying a higher level of reasoning might either be creating a model to solve a problem or alternatively using a familiar model to generalise about a new problem.

Verhage and De Lange (1997), also stress the importance of context if the process of mathematisation is to take place. They state that problems can be divided according to three levels of context. The first level is context-free, which means the task has not been set within any context. In the second level a context may be provided but is regarded as 'camouflage context' or where the context is provided to give a relevance to the task but is not an essential element of the task. The third level requires the context to be authentic and an essential and relevant part of the task at hand. Verhage and De Lange maintain that authentic contexts are necessary if students are to engage in higher order thinking. The pyramid structure demonstrates the multi-faceted nature of problems and tasks. For example, a task may require a lower level of reasoning but require complex mathematical skills. Conversely, a task may require a high level of reasoning but require only medium complexity mathematical skills.

Often when teachers approach the topic of time series, they initially try to select data sets whose contexts are familiar to students. Within a familiar context the level of reasoning demanded is lower than when a context is unfamiliar. Similarly the complexity of the task will increase if the context is unfamiliar to the student. For time series the NCEA assessment is an extended piece of work which suggests that students will be able to be assessed for higher level thinking. Thus students have the opportunity to demonstrate a critical attitude and an ability to interpret and reason from the data at a high or excellence level.

SOLO Taxonomy

The Structure of the Observed Learning Outcome (SOLO) taxonomy was developed by Kevin Collis and John Biggs (1982) and has grown in popularity. SOLO taxonomy aims to assess the structural level of a student’s learning and is based on cognitive theory. It allows student learning to be assessed in terms of its complexity and quality rather than how many parts of an assessment they got right or wrong. For teachers it can help them to assess where students are placed in a hierarchical structure of learning based on students’ responses. This in turn can provide guidance as to what future instruction is required in order to assist students in attaining the next level of learning.

Creation of the taxonomy began with Piaget’s stages of development in order to establish a categorisation system for student responses. Pegg (2005) gives five modes of functioning with approximate age ranges associated with each one:

1. *Sensori-motor* (from birth). Physical co-ordination skills often in response to an external stimulus.

2. *Ikonic* (from around 18 months). For example, the development of language.

3. *Concrete-symbolic* (from around 6 years). Involves use of written language, numbers, symbols, signs and charts.

4. *Formal* (from around 16 years). The ability to hypothesise and refine structures.

5. *Post Formal* (from around 20 years). The level usually associated with research as it involves redefining and perhaps extending existing boundaries or ideas.

Within each mode identified above learners can operate at one of five different levels. Five structural levels are identified:

1. *Pre-structural.* A student may focus on one particular aspect, perhaps as a result of some prior knowledge, but effort is often irrelevant to the content of the task.

2. *Uni-structural*. A student begins to focus on one relevant aspect of the task.

3. *Multi-structural.* A student is able to work on several aspects of the task, but is unable to relate or integrate them.

4. *Relational.* A student can work on several aspects of the task and integrate them to produce a coherent and meaningful solution.

5. *Extended abstract.* A student can generalise the task to an, as yet, untaught application.

The SOLO taxonomy is illustrated in Figure 2 (Pegg, 1992). The diagram portrays a fairly linear progression from one level to the next which anyone with extensive teaching experience is likely to question. However, Romberg, Zarinnia and Collis (1990, p. 21), describe a process whereby "modes do not successively replace each other…. but as each develops, it is added to its predecessor (which itself continues to develop) and, thus the modal repertoire of the mature adult" is created.

Image removed from final copy due to copyright. Please see citation below.

Figure 2 Modes, learning cycles and forms of knowledge of the SOLO taxonomy

Reprinted from Pegg, J. (1992). Assessing students' understanding at the primary and secondary level in the mathematical sciences. In M.Stephens & J.Izard (Eds.), *Reshaping Assessment Practice: Assessment in the Mathematical Sciences Under Challenge* (pp. 368-385). Copyright 1992 by Australian Council of Educational Research. Reprinted with permission.

Time series analysis, which this research focuses on, is taught in Year 13, and therefore the mode most likely to be involved will be formal, but within that any of the five structural levels might be observed. At the uni-structural level, the student of time series will be capable of carrying out a particular task such as identification of peaks and troughs, but would make no attempt to link this information with the context of the data or to interpret why the peaks or troughs have occurred. At the multi-structural level, the student of time series may be able to identify peaks and troughs, smooth data, identify unusual values and perhaps describe long and short term trends, but they would be unable to relate these properties in order to produce an holistic picture of the time series. At a relational level, not only will the student be able to carry out the tasks mentioned previously but would also be able to relate characteristics of the time series to its context and attempt to interpret those features and offer possible explanations for them. At an extended abstract level, students would be able to analyse time series data from unfamiliar contexts and, through further study and research, be able to provide a coherent interpretation of the time series' characteristics.

4.2.2 Framework characterising the development of mathematical reasoning

The Pirie-Kieren framework (Figure 3) was selected for consideration because it is well-established as a model to describe students' growth of mathematical reasoning and therefore may provide some insights into the development of the time series framework.

Pirie-Kieren Theory

The Pirie-Kieren theory of the development of mathematical understanding was first described in 1989 and has since become a well-established model (Martin, 2008; Pirie & Kieren, 1994). A fundamental component of this model is that understanding in mathematics is a dynamic process which involves continual movement between different levels of reasoning and understanding. This dynamic process is not seen as a linear one but one that may involve re-visiting and re-examining previous levels of understanding in order to progress to a higher level of understanding. The process of re-examination is termed *folding-back* within the Pirie-Kieren theory. The framework of the Pirie-Kieren model comprises eight layers:

1. *Primitive knowing*. Prior knowledge which can be utilised for the topic currently being studied. It does not necessarily mean low level knowledge.

2. *Image making.* The student will be involved in activities aimed at assisting them to understand a new mathematical topic or idea. These activities may involve visual images but verbal images or actions may also be indicative of this level.

3. *Image having.* The student's understanding allows them to undertake an activity without the visual, verbal or action image for support. A student at this stage has a mental image of what is required.

4. *Property noticing.* The student is able to investigate and explore the mental images they possess and begin to make connections between them, identify common properties or differences and articulate what they have discovered. The ability to reflect and critique their own work is a characteristic of this level.

5. *Formalising.* The student is able to generalise about properties and work with the concept as a formal object.

6. *Observing.* The student identifies connections between formal statements of a concept and defines his/her ideas as algorithms or proofs.

7. *Structuring.* The student examines his/her formal observations as a theory and may begin to construct logical arguments in the form of a proof.

8. *Inventising.* The student has full understanding of the concept and can pose questions about that concept that may lead to investigation and understanding of new concepts.

Image removed from final copy due to copyright. Please see citation below.

Figure 3 The Pirie-Kieren model of the growth of mathematical understanding

Reprinted from Pirie,S. & Kieren, T. (1994, March). Growth in mathematical understanding: How can we characterise it and how can we represent it? *Educational Studies in Mathematics, 26*(2-3), 165-190. Copyright 1994 by Springer. Reprinted with permission.

Although evidence of all of these levels may be hard to find in a time series written assessment task, there are some aspects of this framework which deserve to be highlighted. The level identified as property noticing is probably one of the most important for time series analysis as it is when students can begin to articulate properties and differences between time series. The final level of mathematical understanding, inventising, which the authors mention is not often reached includes the ability to pose questions that may lead to further investigation. This ability is particularly important and characteristic of statistical reasoning as the next section illustrates.

4.2.3 Frameworks characterising the development and dimensions of statistical reasoning

In this section a framework developed for statistical thinking in empirical enquiry will be examined and two frameworks that were initially developed for assessing the interpretation of data and data displays, sometimes termed graphicacy, will also be discussed. These latter two frameworks are included in this section since interpretation of data displays is an integral part of statistics in general and in time series in particular.

A framework for statistical thinking in empirical enquiry

The theoretical framework discussed in the previous section related to mathematical thinking. In this section a framework that specifically considers the characterisation of statistical rather than mathematical thinking is discussed. Although some might maintain that development of mathematical and statistical thinking is similar if not the same, Resnick, (1987) maintains that each discipline has its own characteristic ways of reasoning. In statistics for example, the integration of context is paramount; yet only the RME framework devoted any attention to the importance of context. The following framework developed by Wild and Pfannkuch (1999) resulted from a literature review in conjunction with in-depth interviews with students of statistics and practising statisticians. The framework consists of four dimensions illustrated in Figure 4.

Image removed from final copy due to copyright. Please see citation below.

Figure 4 Framework for statistical thinking in empirical enquiry

Reprinted from C. Wild, & M. Pfannkuch (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), p 226. Copyright 1999 by John Wiley and Sons. Reprinted with permission.

In seeking to develop these dimensions some might argue that although the mechanism of the Problem, Plan, Data, Analysis, Conclusion (PPDAC) cycle may be taught, dispositions may not be taught or promoted since they could be considered as innate characteristics of the student. Wild and Pfannkuch believe that as experience in statistical thinking is gained students become aware of representations and data that have solid foundations and those that do not. Through experience students can accumulate a healthy dose of scepticism prompting questions about the reliability of data sources and collection processes. Hence they develop the characteristics listed under dispositions as they become more proficient in statistical enquiry techniques.

All four dimensions have a part to play in the analysis of time series but perhaps some aspects play a more prominent role. For example in dimension 2 – types of thinking – transnumeration, consideration of variation, integration of statistical and contextual knowledge are particularly important. Transnumeration is the process by which students view a data set through a number of different representations in order to gain deeper understanding. In time series analysis, representations that highlight the components of a time series greatly assist that understanding. Integration of statistical and contextual knowledge is also essential if the variation in a time series is to be meaningfully interpreted. Reasoning with statistical models is also important with time series. Students must comprehend the underlying components of a time series in order to tell the story of the time series effectively. Dimension 3 represents the interrogative cycle. In time series analysis the continual posing of 'What if' or 'I wonder' style questions by students will help them interpret a data set. Each question posed and subsequently answered may give rise to a further question or questions. A good student of statistics definitely displays several if not all of the traits identified in Dimension 4 – dispositions. Although some students may naturally have a more inquisitive or sceptical nature, this does not mean that a student cannot learn to investigate data sets without such an attitude of mind. This style of investigative approach has the potential to be replicated by all students of statistics.

Frameworks for assessing interpretation of data and data displays

The theoretical perspectives discussed in the previous sections can be used to model the development of any area of mathematical and statistical understanding. However, an important component of the topic of time series analysis is the interpretation of data and graphical displays. A considerable amount of research has been conducted on how students develop an ability to interpret graphs and two frameworks have been created to help understand how this process takes place.

Firstly a framework for the different levels of understanding students have about data developed by Konold, Higgins, Russell, and Khahil (2015) will be considered. These researchers analysed a large quantity of student data to produce four general perspectives that students use when working with data. The perspective of the student can influence the type of questions they are able to pose and their ability to make inferences from the data. The four perspectives are:

1. *Data as pointers.* The data serves to help students recall associated facts or memories rather than information limited to the observed event.

2. *Data as a case value.* These students do not focus on the data as a whole but rather on individual values within the whole data set – particularly those values that might relate to them or to others that they know.

3. *Data as a classifier.* These students are very focussed on the mode or the 'winning' outcome, but have little notion of the overall distribution of the data. The preferred method of display is usually one where individual frequencies can be identified, rather than a display that provides an overview of the distribution.

4. *Data as an aggregate.* Students who can view the data as an aggregate can focus on the entire distribution of the data, they are not focussed on individual values, They can calculate with summary statistics such as measure of central tendency or measures of spread and can use percentages to highlight proportions lying between particular values. A major barrier to viewing data as an aggregate is the inability of students to see that different observations can belong to a single group; a collection of values with its own properties (Stigler, 1999).

In time series analysis students must be able to focus on the complete data set if they are to recognise key characteristics of the time series such as seasonal variation and overall trend. Thus students must be able to view *data as an aggregate,* in Konold's classification scheme, that is, take a global view of the data. However, they also need to take a local view of the data by focussing on individual cases such as outliers or clusters that may need further investigation. Students must be able to differentiate between small and large deviations from overall long term trends, with only larger deviations warranting further investigation thereby avoiding obfuscation of the overall picture by minor deviations.

As mentioned previously students of time series not only have to be data handlers; another important component is their ability to comprehend data presented in a graphical form. Friel, Curcio and Bright (2001, p. 132) define graph comprehension as "the ability of graph readers to derive meaning from graphs created by themselves or by others". They categorised two main classes of graphs; those used primarily for analysis and those used primarily for communication.

Curcio (1987), developed a framework for graph comprehension comprising of three levels:

1. *Reading the graph*. Extracting individual pieces of data or information from the graph. To do this the student must understand the scale and the units of measurement.

2. *Reading within the graph*. Involves some degree of interpretation and integration of information that is presented in the graph. For example, this could involve using several data values to perform a meaningful calculation.

3. *Reading beyond the graph*. Requires some inference based on the graph, for example, interpolation, extrapolation or prediction. At this level a student would be able to pose and analyse exploratory questions about the data.

Shaughnessy (2007) argues that a fourth level of graph comprehension exists which sits beyond the three levels defined by Curcio.

4. *Reading behind the graph*. Statistics always has a context and the unique role of statistics as a discipline is identifying the connections between the context and the graph; that is, what is behind the graph (Wild & Pfannkuch, 1999). Reading behind a graph may involve further investigation of patterns in the data and features in the data, suggesting possible causes for them.

Students of time series analysis should display all of these four levels. A student who exhibits higher order thinking skills in this topic must engage with the context of the time series and consequently will be reading behind the graph according to the classification levels given above.

4.3 Summary of key components of frameworks

Key components from each framework are summarised in Table 3. Apart from the statistical thinking framework, all of the other frameworks are hierarchical in nature which tend to suggest a linear progression in the development of student reasoning from one level to the next. However, these frameworks stress that progression is not necessarily linear but may involve revisiting lower levels of reasoning in order to progress to a higher level of reasoning. This type of development, the Pirie-Kiren framework terms *folding back*. Thus anyone involved in instruction should be aware that successful progress may involve re-visiting areas several times before the next level of reasoning can be mastered.

However, one fairly consistent theme across these hierarchical frameworks concerns higher order thinking skills. The higher levels of reasoning in all frameworks require a student to use their mathematical or statistical knowledge in some unfamiliar way. This unfamiliarity might be to use a skill in a different application, to extend that skill independently or particularly in statistical topics to use that skill to integrate the student's statistical and contextual knowledge with the goal of improving interpretation. Lower order thinking skills across the frameworks include performing calculations according to learned algorithms.

A particular feature of the framework for statistical thinking in empirical enquiry is the cyclical nature of statistical investigations. An initial problem is posed but after some preliminary analysis this may raise further questions. Similar cycles of enquiry may also occur in mathematics but they are always present in statistical investigations. Cycles present in the mathematical reasoning frameworks are described as revisiting lower levels of reasoning to consolidate knowledge or to re-align previous knowledge with new knowledge. Thus the cycles in the mathematical reasoning frameworks are a *mechanism* by which a student can develop higher levels of reasoning. Statistical cycles are somewhat different in nature; a statistical cycle does not characterise a level of statistical reasoning but rather a *mechanism* for empirical enquiry. Repetition of statistical cycles may, but do not of themselves, generate higher levels of statistical thinking.

Table 3 Key components of theoretical frameworks

Framework name	Primary focus of framework	Authors	Cognitive levels	Key ideas used in synthesised framework
De Lange's Pyramid	Assessment analysis	Verhage & De Lange (1997)	Levels of reasoning – low, middle, higher Context – context free, camouflage & authentic Complexity – simple, middle, complex	Context is crucial in any statistical investigation. Higher levels of thinking cannot be accessed without integration of context.
SOLO taxonomy	Assessment analysis mathematical reasoning	Collis & Biggs (1982)	Five modes: Sensori – Motor, ikonic, concrete-symbolic, formal & post-formal Within each mode: Pre-structural, uni-structural, multi-structural, relational & Extended abstract	Students in year 13 will be at formal level but will be drawing on ikonic and concrete-symbolic modes.
Pirie- Kieren	Mathematical reasoning	Pirie & Kieren (1994)	Primitive knowing Image making Image having Property noticing Formalising Observing Structuring Inventising	Property noticing is an important stage of reasoning in time series analysis.
Statistical thinking in empirical enquiry	Statistical reasoning	Wild & Pfannkuch (1999)	Investigative cycle Interrogative cycle Types of thinking – including transnumeration, consideration of variation, reasoning from models and integration of statistical and contextual knowledge Dispositions	All four dimensions are present in the analysis of time series.
	Graphicacy	Konold et al. (2015)	Data as a pointer Data as a case value Data as a classifier Data as an aggregate	To interpret time series students must be at the level of viewing data as an aggregate.
	Graphicacy	Curcio (1987) Shaughnessy (2007)	Reading the graph Reading within the graph Reading beyond the graph Reading behind the graph	Higher order thinking skills are evident at the 4th level. To comprehend graphs students should be proficient at all four levels.

4.4 Framework synthesising characteristics of types and levels of reasoning for time series

Table 4 shows a framework synthesising characteristics of types and levels of reasoning for time series, which were constructed from the frameworks summarised in Table 3 and described in Section 4.2. This framework will be used to assess students' responses for the types and levels of reasoning they are using in time series investigations.

The following hierarchy of six levels of reasoning are proposed for the time series framework:

1. *Vertical reasoning*	Reading individual items of data from a time series graph.
2. *Horizontal reasoning*	Reading across the whole time series; identification of basic features of the time series.
3. *Procedural reasoning*	Performing calculations using the time series data.
4. *Extended procedural reasoning*	Using calculations to predict beyond the current domain or to interpolate for missing values.
5. *Interpretive reasoning*	Interpreting key components and features of a time series through integration of statistical and contextual knowledge.
6. *Interrogative reasoning*	Critical analysis of statistical model and interpretation offered. Ability to use skills in an unfamiliar context. Multiple repetitions of enquiry cycle – each time trying to extend and refine interpretation of time series.

In the framework some of these levels have been broken down into further detail so that when student work is examined, differences within a level as well as between levels can be identified (see codes in Table 4). Categories within overall levels have been numbered as sub-categories, (e.g. level 2 has sub-categories 2.1, 2.2 & 2.3) but in fact no hierarchy is intended within levels only between levels. The sub-category coding simply represents different ways of demonstrating a particular level of reasoning. It is anticipated that in a

Level 3 Achievement Standard, the level of reasoning required, particularly at the Excellence level should reach the higher levels of reasoning described in this framework.

Each of the proposed six levels of reasoning for the framework will now be discussed in terms of how the reasoning will be explicated by the students in their time series investigations. Note that Table 4 gives descriptions for each of these levels as well as the theoretical framework literature on which each descriptor was formulated.

Table 4 Synthesised framework of types and levels of reasoning with time series

Level	Types of Reasoning	Characteristics	Codes	Descriptors for Time Series	Theoretical Frameworks
1	Vertical Reasoning	Ability to read and understand scale, units and basic graph construction. Context may be present but not critical to task.	1	Individual data values read from a graph.	Konold - Data as a pointer, data as a case value Curcio - Reading the graph SOLO- concrete-symbolic, unistructural RME - context free/camouflage, low level reasoning, simple complexity
2	Horizontal Reasoning	Looks across the data not just individual data values. Context may be present but not critical to performance of tasks.	2.1	Reads across data to assess overall trend direction	Konold - Data as an aggregate
			2.2	Identifies when peaks and troughs occur	Curcio - Reading within the graph SOLO- concrete-symbolic/formal, unistructural/multi-structural
			2.3	Identifies any unusual values	RME - context free/camouflage, low level reasoning, simple complexity
3	Procedural Reasoning	Uses mathematical procedures to calculate properties of data set or to examine features of data set. Context may be provided but not critical to performance of procedure.	3.1	Calculates smoothed data	RME - context free/camouflage, low/middle level reasoning, middle complexity
			3.2	Calculates seasonal effects	SOLO- formal, unistructural/multi-structural
			3.3	Finds equation of line of best fit	SOLO- formal, unistructural/multi-structural
			3.4	Estimates trend using line of best fit	
			3.5	Calculates seasonally adjusted data	RME - context free/camouflage, low/middle level reasoning, middle complexity
4	Extended Procedural Reasoning	Uses a mathematical model to predict beyond the current domain of the data set or to interpolate within the given domain. Context may be provided but not critical to performance of procedure.	4.1	Uses mathematical model to make predictions beyond current data set	Curcio - Reading beyond the graph. SOLO - formal, multi-structural
			4.2	Discusses validity of predictions	RME - context free/camouflage, middle level reasoning, middle complexity
			4.3	Mathematical model used to find missing values within domain of current data set	RME - context free/camouflage, low/middle level reasoning, middle complexity

5	Interpretive Reasoning	Ability to interpret implications and characteristics of a mathematical model. Transnumeration - examination of different representations of the data to aid understanding. Context must be present and any interpretation provided must refer to it.	5.1	Describes contribution of various components of time series and attempt possible interpretation in context	Shaughnessy - Reading behind the graph Pirie-Kieren – Property Noticing
			5.2	Identifies variation in trend and data over time and provides of possible interpretation in context	RME - authentic context, middle/high level reasoning, middle/high complexity
			5.3	Discusses model limitations in context	SOLO - formal, relational
			5.4	Considers other factors to assist in seeking for explanations of variation in time series	Wild & Pfannkuch - Investigative Cycle (PPDAC), Interrogative Cycle, Types of thinking including transnumeration, consideration of variation and reasoning with a statistical model
6	Interrogative Reasoning	Seeks information to assist in interpretation or that may confirm or refute own findings. Ability to critique own analysis and to raise further questions. Has a healthy dose of scepticism. Integration of contextual and statistical knowledge essential.	6.1	Examines other research or findings to confirm or refute own findings.	Wild & Pfannkuch - Investigative Cycle (PPDAC), Interrogative Cycle, Types of thinking including transnumeration, consideration of variation and reasoning with a statistical model
			6.2	Considers in detail other factors to assist in seeking for explanations of variation in time series	Wild & Pfannkuch - Dispositions including scepticism, curiosity and perseverance
			6.3	Acknowledges that their analysis is not exhaustive and that other interpretation may be valid.	SOLO - formal/post formal, relational/ extended abstract
			6.4	Integrates statistical and contextual knowledge	RME - authentic context, high level reasoning, high complexity

Levels 1 and 2 - Vertical and horizontal reasoning

The first level of reasoning in time series is when a student can view a time series presented graphically or in a table and identify particular values from that graph or table. Such a student would be able to answer questions like:

"In what year(s) was the unemployment figure over 500,000?"

"What is the lowest Dunedin temperature since weather data was first collected?"

However, once a student has progressed to horizontal reasoning they are able to look not only at individual cases of data but also across the whole time series and begin to identify key features of the time series, such as seasonal cycles, overall trend and any unusual values. A student at this level of reasoning can respond to questions such as:

"Describe the seasonal cycle of sales of ice-cream in Christchurch".

"Describe the trend in in sales of smart phones since 2005".

"Identify any unusual values in visitor numbers to New Zealand since 2000".

Level 3 - Procedural reasoning

For the procedural level of reasoning students are able to perform certain calculations with the time series. This may involve the production of some sort of smoothed data, calculation of average seasonal effects, or finding a model for the overall trend. Although a student may be able to produce a smoothed time series using a formula or some software, little further analysis is evident. Procedural reasoning involves students in number crunching or use of technology but with no attempt to decipher any meaning or improve their understanding of the time series. In order to perform these types of calculations contextual knowledge is not essential as no attempt at interpretation is made.

Level 4 – Extended procedural reasoning

The difference between procedural and extended procedural reasoning is simply that calculations are performed for values outside of the current domain of the time series. This might involve using a model to produce predictions beyond the current time frame or to interpolate within the current domain for missing values. Again contextual knowledge is not essential in the performance of these tasks although of course it is essential if it is to be done in a meaningful way. For example, a student may model the trend with a parabolic curve, which might provide a reasonable fit for the data in the given domain, but knowledge of the context might suggest that extension of this parabola beyond the current domain would provide completely unrealistic predictions.

Level 5 - Interpretive reasoning

Interpretive reasoning represents a considerable advance over the procedural reasoning levels. In terms of instruction this level of reasoning would usually be introduced by analysing a time series example whose context was familiar to students and about which they were engaged and had some prior knowledge. Such an example might be described, according to De Lange's pyramid model, as familiar authentic context, middle level reasoning and middle level complexity. However, if the context is unfamiliar to students and they have no prior knowledge of the context, the example may be described, according to the same classification system, as unfamiliar authentic context, higher level reasoning and higher level complexity. In other words familiarity with the context is a huge advantage in statistical analysis. Practising statisticians, however, need to be able to work in unfamiliar contexts too and need to be able to assimilate knowledge about that context quickly.

Students at this level will be able to interpret different representations of the time series in order to promote their understanding of the underlying characteristics of that time series – this is the notion of transnumeration identified in the statistical reasoning framework.

Students at this level of reasoning will be able to tackle questions such as:

"Discuss and interpret the reasons behind the seasonal cycle of power usage in NZ households".

"Discuss and interpret trends in the number of new building permits approved in Christchurch during the last 10 years".

Level 6 - Interrogative reasoning

Interrogative reasoning is the highest level of reasoning identified. At this level students will investigate other research or articles to discover whether their interpretation of the time series can be confirmed or refuted. They will investigate similar or related time series to assist in the explanation of the variation in the time series under consideration. They will be able to integrate their statistical and contextual knowledge in familiar and unfamiliar contexts. Most importantly they will be able to critique their own analysis, highlight any assumptions made and acknowledge the possibility of alternative interpretations. They will continue to pose questions and answer them in their quest to gain deeper understanding of their data set.

Students at this level will also utilise transnumeration, which may involve changing the current representation of the data, re-categorisation of variables or splitting data into groups in order to facilitate telling the story of the data.

Students at this level of reasoning will deal competently with questions such as:

"Discuss and interpret trends in two sectors of retail spending. Compare and contrast the two sectors and relate variation in these sectors to other factors that you consider important".

"Discuss and interpret trends in migration to NZ by ethnic group. Compare and contrast at least two groups that have different migration patterns and research possible reasons behind these different patterns".

4.5 Conclusion

As mentioned at the outset of this chapter the purpose of producing this framework for reasoning from and about time series analysis was to answer the research questions, but it may also help inform instruction and assessment of time series. Many teachers would state that their instruction goal is to increase the level of reasoning of their students. However, studies of assessment in particular over several decades show that many teacher-authored tests require only simple recall of information, demanding only low levels of reasoning (Marso and Pigge, 1993, cited in Brookhart, 2010). Furthermore, when teachers are surveyed about what level of reasoning they think they are assessing, they frequently report their assessments are aimed at higher order levels of reasoning (McMillan,2001, cited in Brookhart, 2010). Brookhart (2010) suggests that the reason so many assessments focus on testing low levels of reasoning is that these types of assessment are the easiest to write. However, it is not only in assessment that teachers often focus on the lower levels of reasoning. Classroom discussion questions posed without preparation often require only low levels of reasoning – factual questions for example requiring recall but no higher level of reasoning. With preparation classroom discussion questions can be reviewed to ensure that higher levels of reasoning are also developed.

Anderson and Krathwohl (2001) define the notion of higher order thinking as a combination of learning for recall and learning for transfer. It is the 'transfer' element of this definition that I believe is the most significant. According to their definition, a student will be able to use their knowledge and skill set developed during learning in unfamiliar contexts. They state that higher order thinking skills enable a student to relate their learning to areas not only in unfamiliar contexts but also those beyond their learning experience. I believe these characteristics of higher order thinking are confirmed and reflected in levels 5 and 6 of the framework shown in Table 4.

CHAPTER FIVE

Research methodology

5.1 Introduction

A variety of research methods exist for mathematics education research that can be utilised to assist in answering a wide range of research questions. Kelly and Lesh (2000) maintain that research questions should not be restricted to those for which research methods currently exist, but rather that the research question is paramount and methods should be adapted, if necessary, in order to seek answers. Some may argue that early consideration of methodology is vital as a poorly constructed survey, for example, is unlikely to yield any useful information. I think there is a middle ground – research questions should not initially be constrained by available methodology as it is important that researchers have the freedom to pose research questions without restrictions. However, in posing a research question the researcher inevitably must consider possible methods of data collection even if specific details are yet to be finalised. This research is an attempt to confirm or refute anecdotal evidence that the recent curriculum changes in the teaching and assessment of time series analysis has improved student outcomes. Although, this research question was posed, initially, without any thought of methodological issues; some thought about possible methods of data collection and precisely what data could be collected in a limited time-frame were considered but not finalised from the outset.

5.2 Overview of research methods

Choice of research method is primarily determined by the type of data a researcher is intending to collect. There are two main types of data:

> *Quantitative data*. Data that is measured or counted, thus the data will often be numerical in nature.
>
> *Qualitative data*. In simple terms qualitative data is any non-numerical data, but within that there are several different types. For example, interview transcripts and observational notes would both be considered examples of qualitative data.

Some researchers may seek a mixture of quantitative and qualitative data, if it appears appropriate for their particular research project. There are no rules as to which should be used or in what ratios as each piece of research will have different needs. However, the instruments of data collection will vary according to the type of data being collected.

Qualitative and quantitative data are often described as complementary to each other, but others (eg., Corbin and Strauss, 2007) suggest that it is the interplay between the two types of data that researchers need to be aware of. An initial collection of qualitative data, for example, can inform what quantitative data should be collected. This feedback between the two data types may have a circular flow, so that data collection gradually evolves during the research process.

Many handbooks have been written which describe the range of education research methods available. Descriptions of some of these methods are provided below. This is not an exhaustive list of methods but does include the more commonly used methods. Cohen, Manion and Morrison (2013) do not suggest that any one style in particular should be slavishly adhered to, rather that styles are matched appropriately to the research question and mixed as required. A variety of different research methods mentioned by Cohen et al. and other authors are discussed below.

Naturalistic and ethnographic research. This involves a style of research that describes events or activities within their cultural context. The researcher attempts to record data on the perceptions of members of the group under observation (Hitchcock and Hughes, 1995, cited in Cohen et al., 2013). Ethnography is neither objective nor subjective but rather an interpretation of group participants' interpretation of their situation. As such this method if used by a different researcher may reveal different characteristics about the group in question as perceptions may be interpreted in a different way. Ethnographic research can have two forms – ethnology and ethnohistory. Both methods are interested in studies of cultural groups within their cultural setting, but each takes on a slightly different perspective. Ethnological studies will consider a number of studies about the same cultural group in order to identify patterns of behaviour and social norms. For example, researchers have found that most societies display some sort of extended family relationship but that pattern varies considerably between different cultures. Ethnological research would seek to identify similarities and differences. An ethnohistorical research project, on the other hand, will usually focus on one particular cultural group and through examination of historical documentation and discourse with elders of the group seek to build up a picture of changes experienced by this one particular group. An ethnohistorical research project is often conducted to inform a larger ethnographical study (Johnson & Christensen, 2004).

Historical and documentary research. Borg (1963, cited in Cohen et al., 2013) describes this style as the systematic and objective location, evaluation and synthesis of evidence in order to establish facts and draw conclusions about past events. Historical research may be simply descriptive or may attempt an interpretation of events. Historical research may also access oral sources, such as oral histories (Gardner, 2003). Historical research cannot be strictly regarded as scientific as it does not provide the opportunity to repeat an observation or an experiment. However, this method still satisfies many other principles of scientific research (Borg, 1963, cited in Cohen et al., 2013). Documentary research is also sometimes termed content analysis. An advantage of this type of research is that the researcher can investigate past patterns and trends and is not restricted to collecting only current data. Accessibility of historical documents via electronic media can be straightforward and hence this type of research can be attractive to researchers with limited time and money. Validity of this type of research is problematical. With such a vast range of historical information available the researcher has to be selective in the documents they analyse and as such the validity of the research requires careful scrutiny (Wallen & Fraenkel, 2001).

Surveys. There are two types of surveys - longitudinal and cross-sectional. A longitudinal survey collects data from the same group of subjects often over a considerable period of time. There are three types of longitudinal surveys – a trend survey, a cohort survey and a panel survey. A trend survey involves selecting a number of different samples from the same population over time. A cohort survey collects information from the same group of people over time and a panel survey focuses on a particular group of individuals and tracks their changes over time. Thus a cohort study might be of all those born at a particular hospital in a particular month, whereas a panel study is more targeted, for example, looking at graduates from a particular course in a particular year (Wallen & Fraenkel, 2001). A cross-sectional survey provides a snap-shot of a particular group at one particular time. Large scale surveys are a common choice with researchers seeking to generalise from their results. Small scale surveys are also possible but results are unlikely to be generalizable (Punch, 2009). If a survey is conducted for a complete population, it is known as a census.

Case studies. Hitchcock and Hughes (1995, cited in Cohen et al., 2013), maintain that case studies are defined not so much by the methodology employed but more by the subjects of the enquiry. They suggest that this approach is often used when the researcher cannot control or manipulate behaviours. A number of factors, not under the control of the researcher, may define the case study such as geographical, organisational or institutional. Thus a case study might focus on a school cluster in a particular region, an organisation such as a local mathematics association or an institution such as a company.The case study, unlike an historical study may include interviews and direct observation of participants. A case study may collect both qualitative and quantitative data. Case studies are sometimes part of a larger multi-case study research project. In a multi-case study a researcher may compare and contrast individual cases in an attempt to identify any similarities or differences (Johnson et al., 2004).

Experiments. An essential feature of any experiment is that researchers will deliberately control particular factors, then introduce a change and observe or monitor any changes. The factor or variable that is manipulated is called the independent variable; the outcome or outcomes observed after manipulation are known as the dependent variable(s). Often an experiment is conducted using a control group. For example, in education if a new teaching method is to be evaluated, the control class would be taught using existing teaching methods and the experimental class would be taught using the new teaching method. At the end of the teaching unit, the researcher would need to measure the difference between student outcomes under the existing and the new method of teaching (Wallen & Fraenkel, 2001). An attraction of this particular research method is that experiments are the only way to prove cause and effect. However, in educational research the deliberate manipulation of the independent variable can give rise to some serious ethical concerns, as will be discussed in a later section.

Action research. This type of research has been practised since the 1920s (Whyte, 1991; Reason & Bradbury, 2008) and is unique in that is often carried out not solely by researchers but in conjunction with, for example, teachers, administrators or other professionals. The research question itself may have been prompted by the practitioners who have a problem for which they have been unable to formulate a solution. Academic researchers may then be approached to work with the practitioners to design some research that may assist in resolving the specific local problem. Action research may involve several stages, where new insights gained from each stage of the research help to inform and direct the next stage of the research. Often the results of action research are only presented at a local level as many are seeking to solve local problems, but the value of these results may have beneficial implications in other locations or other institutions (Johnson et al, 2004). In education action research is often conducted with a view to changing or improving a particular policy or curriculum (Wellington, 2000).

As mentioned earlier, once a research question has been formulated, the researcher may find that selection of just one of the research methods described above will not allow the question to be answered. Often several methods may be employed in order to answer the research question particularly if qualitative and quantitative data are to be collected. The mix of research methods used in this research is outlined in Section 5.4.

5.3 Overview of research instruments

Research instruments are defined as the methods used to collect data – whether that data be quantitative or qualitative (Murray Thomas, 2003). There are many instruments available, but the five most commonly used are:

Observations. Observation in mathematics or statistics education may involve the researcher simply sitting in a classroom taking notes and recording events according to a prescribed template. Alternatively, it may involve video-taping activities in a classroom for further analysis at a later date.

Interviews. Interviews may be one on one with the researcher or the researcher may interview a group (e.g. several students) at one time. There are a number of different interview strategies that a researcher might employ. They may have a very loosely structured set of interview questions which include open-ended questions and they may not utilise all the questions in every interview. Alternatively, the researcher may have a very structured interview protocol with only closed questions and little if any variation in questions asked between interviewees. Some interview protocols may be response-guided, which means that depending on an interviewee's response certain questions may be skipped or alternate questions may be asked. Finally some researchers may choose a combination of open-ended and closed questions for their interview protocol in order to elicit both quantitative and qualitative data. It is important that the researcher considers the sequence of questions, the length of time to answer and the setting of the interview.

Questionnaires or surveys. A survey or questionnaire can be completed in a number of ways, by post, by telephone, via the internet or a face to face interview. In large scale surveys some degree of sampling is usually required in order to contain costs associated with survey delivery and the time commitment of the researchers. Surveys can be used to collect both factual information and information about perceptions. Murray Thomas (2003) recommends that factual questions precede those regarding opinions or perceptions.

Content analyses. This requires a researcher to search various forms of communication in order to answer their research question. These communications may be letters, memoirs, speeches, newspapers, journals, books or any other form of written communication.

Tests. A researcher may seek to answer their research question by asking participants to respond to a series of questions. For example, in education research students may be required to complete a test before and after the teaching of a particular concept.

Several research instruments were used in this research and they are described in detail in Section 5.4.2 and 5.4.3.

5.4 Research procedure

In the previous sections a number of different research methods and instruments were outlined which are available to the education researcher. In this section the particular procedures used in this research will be described; some problems encountered will be mentioned as well as how these problems were resolved given the circumstances and time frame of the research. Ideal research procedures are discussed as well as reasons why this ideal approach could not be pursued. Modified research procedures arising from compromises made are then described.

5.4.1 Ideal research procedure

Ideally to analyse the levels of reasoning demonstrated by students under each Achievement Standard the same or similar group of students would be taught then assessed under the old and the new standards. In other words the ideal research method would utilise an experimental research model since this could categorically prove whether the new standard had affected student outcomes. Educationally and ethically this is not an option although there may be a few students who repeat Year 13 that may have been taught then assessed under both Achievement Standards in consecutive years. The small group of students who repeat Year 13 were not considered representative of Year 13 students as a whole, so this avenue was not pursued.

Alternatively, but again in an ideal world, in 2012, when the old Achievement Standard, AS 3.1, was still in use, a representative sample of student work could have been selected. However, access to the work of such a sample could not be obtained until 2013, after students had received their grade for the standard to avoid any possible problems regarding grade interference. Permission to use the work of students, who by that time had left school, would be required from the principal, teachers and the students. The researcher would be unable to assist in this work as this could have introduced an element of coercion on participants to respond.

The next step in 2013, when the new standard was first introduced, would be to select a new representative sample of student work. Again access to the work from the sample would not be possible until the year following the assessment, once students had received confirmed results. Permission would also need to be sought from principals, teachers and students who had already left school, posing the same issues as previously highlighted. In 2014, teachers could be approached to gain additional insight into their teaching under both the old and new Achievement Standards. Student work under the old and new standards would be compared and contrasted against the synthesised framework and teachers would be interviewed to gain wider insights into how the teaching of these standards had evolved.

5.4.2 Modified research procedure

However, the reality was that I did not start my Master's thesis until Semester 1, 2014. My thesis was part-time and consequently due for completion over two years; my schedule required me to collect all data for my research in the first year, 2014. Both the old and the new Achievement Standards were internally assessed and it is normal practice for schools to retain copies of internal assessments for the last three years, in case student appeals or moderation issues arise. I knew that exemplars of student work under the old and new standards were currently stored in filing cabinets of schools around the country. The major issue was how I could obtain access to this information and obtain permission to use it for research purposes when the authors of the work were no longer at the school. From each school I would need to request names and addresses of past students whose contact details may have changed. Thus, even if schools had been willing to divulge these contact details there was no guarantee that the details were still current. Alternatively, I could have asked schools to contact their past students on my behalf in order to obtain their permission. However, given the likely workload involved in making the initial contact and then following up non-respondents, it was deemed unlikely that any school would be in a position to undertake this work.

Given my time-frame and the unlikelihood that any school would willingly divulge contact details for past students another approach had to be formulated. Instead of the ideal representative samples of student work under both Achievement Standards, I utilised exemplars of student work available on the internet. Student exemplars are posted online by NZQA as a resource for teachers and are freely available, thus resolving the problem of having to seek permission from the authors. These exemplars would act as a proxy for the ideal representative samples previously outlined. Input from teachers could still be collected from a selection of schools, but these teachers would have no connection to the students who created the exemplars. This solution was not ideal, but it was deemed a reasonable compromise given the circumstances.

5.4.3 Student work exemplars

Historical documents have been used to obtain examples of student work completed under the old AS 3.1 and under the new AS 3.8. These exemplars were made available to teachers by NZQA when each Achievement Standard was introduced into the New Zealand curriculum. Students, who produced this work, cannot be identified. Some of the exemplars were graded according to the NCEA classifications of Not Achieved, Achieved, Merit and Excellence. Several exemplars at each grading level were distributed to give teachers an idea of requirements of each level but also an indication of student work at just over or just under each boundary level. A further five exemplars which were used at NZQA-run work-shops do not have any grading classification as usually the goal of a workshop was to grade the work. Numbers of exemplars at each level and for each standard are given in Table 5. A total of 34 student exemplars were considered.

Table 5 Student work exemplars from both old and new Achievement Standards

Achievement Standard	Unclassified	Not Achieved	Achieved	Merit	Excellence	Total
Old AS 3.1	1	0	4	6	5	16
New AS 3.8	0	3	6	6	3	18
TOTAL	1	3	10	12	8	34

5.4.4 Teacher participants

Initially nine schools were selected from a list of those who had participated in professional development workshops run by the University of Auckland. Schools selected reflected a mix of single sex and mixed schools, decile ratings and geographical locations. This list was used as a sampling framework as it was accessible and represented a selection of schools who demonstrated an interest in the professional development of their staff. An attempt was made to choose schools from a range of sizes, deciles and geographical areas, although formal sampling methods were not employed. As such the sample can be described as a *purposive sample* (Palinkas, Horowitz, Green, Wisdom, Duan and Hoagwood, 2014). Candidates for such a sample are deliberately identified as potentially rich information sources for the main research question. Seven schools in Auckland were selected, one in Christchurch and one in Dunedin. Three single-sex and six mixed schools were selected. There was one private school, the rest were state and all were decile 7 or above. A decile 1 school was invited to participate but declined to be involved. Stage 1 involved approaching the Principals of nine secondary schools with a letter outlining my research and requesting permission to contact his/her teachers of Year 13 statistics, who had taught both the old and new standards, in order to invite them to complete a questionnaire. Only one school Principal did not grant me permission to approach their statistics teachers. During Stage 2 questionnaires were sent out to teachers at these eight secondary schools, and responses were received from all but two of those schools. A total of 18 completed questionnaires were received from 6 different schools. One school only had one teacher who had taught both old and new time series standards. The final stage, involving participants, invited lead teachers of Year 13 statistics at each of the schools from which responses had been received, apart from the school with just one respondent, to participate in an interview. All five teachers invited were interviewed. An overview of this selection process is summarised in Figure 5.

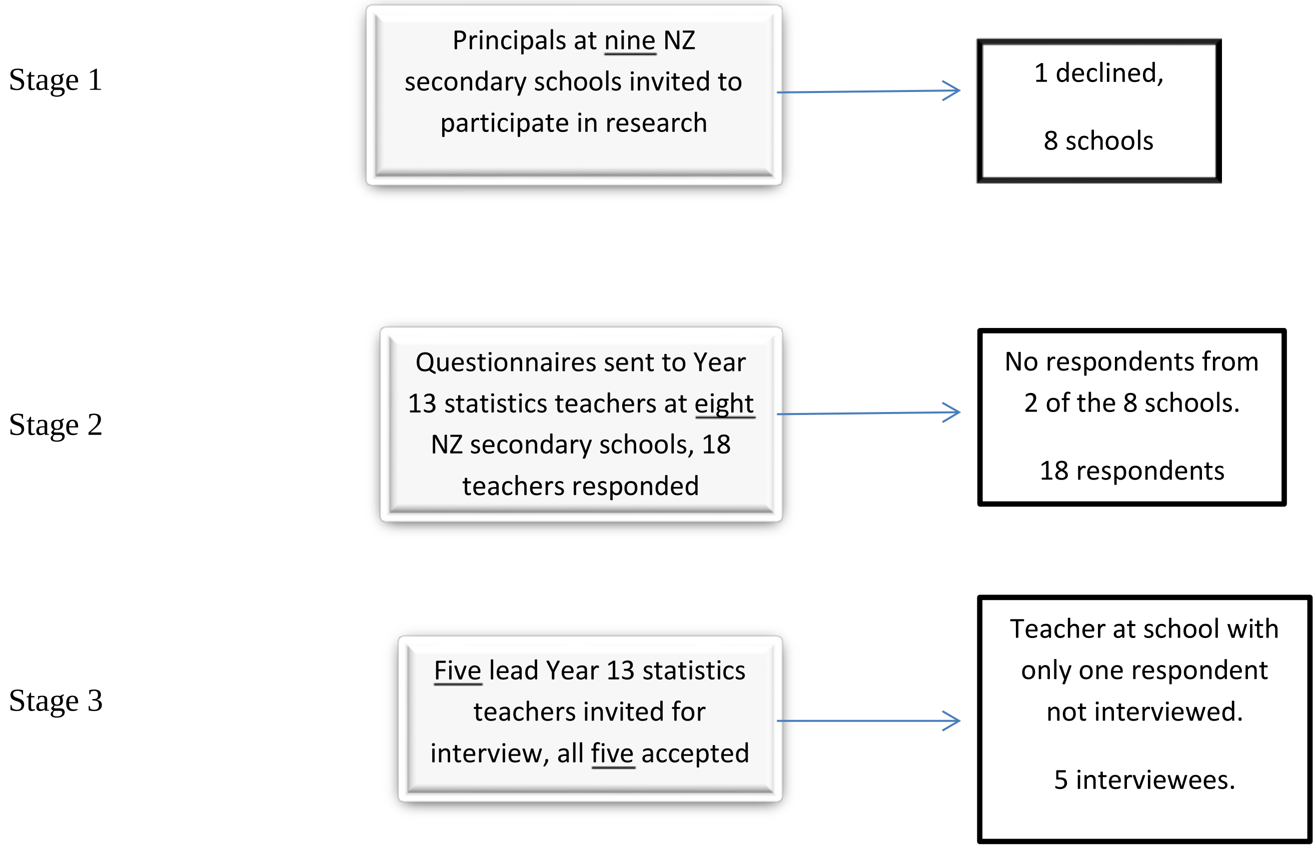


Figure 5 Summary of research selection process

Principals and teachers were all given details of the research and in particular what their involvement might entail via a participation information sheet (Appendix 3). Their consent to participate was obtained before any data was collected. All were offered a copy of the thesis findings at conclusion of the research.

5.4.5 Research instruments

Three different research instruments have been used in this research – content analysis, interviews, and questionnaires. The first research instrument was content analysis, which was used for the exemplars of student work (Table 5). The work was assessed against the synthesised framework (Table 4) developed in Chapter Four.

The second research instrument used was a questionnaire. All teachers at the six schools who had experience of teaching both the old and the new Achievement Standards in time series were invited to complete a postal questionnaire. The questionnaire was designed in six parts.

Part 1 Background information on the school.

Part 2 Background information on the teacher regarding their teaching experience.

Part 3 Details of any professional development undertaken by the teacher prior to the implementation of Achievement Standard 3.8.

Part 4 Comparisons of teaching methods under AS 3.1 as compared to AS 3.8.

Part 5 Details about the assessment process for AS 3.1 and AS 3.8.

Part 6 In this section teachers were invited to note any comments on the questionnaire or any other thoughts about the teaching and learning of time series.

A copy of the questionnaire is given in Appendix 1.

The third research instrument was interviews. The interviews lasted for between 40 and 60 minutes and followed a pre-defined interview protocol. The interview was semi-structured in that although the protocol was largely followed, questions were not answered strictly in order and some later questions were omitted if they had already been answered in earlier responses. This semi-structured format allowed me to pursue any points of particular interest raised by the interviewees, but not covered by the interview protocol. A copy of the interview protocol is given in Appendix 2.

5.5 Quality and validation issues in this research

Most research methods and instruments will have quality and validation issues. Researchers should endeavour to reduce these issues wherever possible or clarify the specific issues if eradication is not an option. A number of quality issues are present in this research, namely quality of documentary data, representativeness, non-response, absence of a pilot questionnaire, interviewer bias, memory recall, interview setting bias and design issues. These are now discussed.

Exemplars of student work were accessed via the internet. Scott (1990), suggests that the quality of such documentary data should be assessed against the following four criteria – *authenticity, credibility, representativeness and clarity of meaning.* Since the exemplars were uploaded to the internet by the NZQA, a reputable organisation, it is safe to assume that the work is authentic and credible. Exemplars of student work are given at a variety of achievement levels for both standards, so in that respect they satisfy the representativeness criterion. Students who completed this work were not identified so any problems regarding clarity of meaning in their responses could not be pursued. Although, the two sets of work can be presumed to be from two different groups of students who had each experienced a different teaching programme, the students were not randomly allocated to either group. This lack of random allocation therefore falls short of the rigour required of a true scientific experiment. As a consequence any generalisation made from this research should be viewed with caution.

A perennial problem with questionnaires is that of *non-response*. I initially contacted the Principals of nine different schools for permission to contact their teachers of Year 13 statistics. All but one Principal gave their permission. Questionnaires were then sent to all teachers of Year 13 statistics at these eight secondary schools, but responses from teachers at only five of the secondary schools were received. In order to try and reduce non-response I included pre-paid and addressed envelopes for responses and I sent several follow up reminder emails. A total of 18 questionnaires were completed and returned out of 29 questionnaires distributed, a response rate of 62%. Some Year 13 teachers who received the questionnaire had not taught both AS 3.1 and AS 3.8 and did not complete the questionnaire as a result. However, I had no way of knowing which teachers did or did not have this experience so chose to send questionnaires to all Year 13 teachers of statistics at each school.

Due to time restrictions there was *no pilot questionnaire* which might have resulted in modification of some questions. Several teachers reported difficulties in allocating amounts of time spent teaching particular time series concepts. However, it is still interesting to discover which items were taught under both Achievement Standards and which under only one – either the old or the new, even if time estimates are unavailable. Other questions and overall length of the questionnaire might have been modified following a pilot. This may have improved the quality of responses and avoided questionnaire fatigue common with lengthy questionnaires. Questionnaire fatigue can mean that some questions are left unanswered or are answered quickly in order to finish (Krosnick and Presser, 2010).

At each school from which completed questionnaires were received the lead Year 13 statistics teacher was invited to participate in an interview. All those invited accepted and interviews were conducted in person or via Skype during November 2014. All interviews were recorded and later transcribed for analysis. As with any interview, responses can be affected by *interviewer bias*, i.e. respondents may feel an obligation to give the answer they think the interviewer is expecting rather than what they really think. Interviewer bias can be a particular problem if opinions on sensitive issues are being sought (Clegg, 2007). Questions posed during the interview were not of a particular sensitive nature, but participants' responses may have been affected by interviewer bias. The interviewer was known to many teachers of Year 13 statistics as a presenter of workshops about the new time series Achievement Standard and this could have affected responses. Some of the questions posed required interviewees to remember how they taught time series analysis in previous years and as such responses relied on *memory recall*. Accuracy of memory recall can also be a problem for data collected by interview, particularly, recall about perceptions which are almost impossible to validate by other means. All interviewees were offered the opportunity to view and amend the transcript of their interview within a pre-determined time frame.

To avoid *interview setting bias*, all interviews were conducted by me at the place of work of the interviewee. Interviewing participants in a familiar surrounding can help the interviewee feel relaxed and able to focus on the interview rather than been distracted by an unfamiliar location (Kvale, 1996). Recording of the interviews was done using Sound Recorder on a laptop, thus interviewees could speak normally without feeling apprehensive about the technology used.

Both quantitative and qualitative data were collected using the research instruments. Such a deliberate attempt to collect complementary data about the same topic is sometimes referred to as a *triangulation design*. Triangulation can help to validate or give a more holistic description of the topic in question, and this was the motive in this research (Stake, 2010). There are a number of different triangulation designs including embedded, explanatory and exploratory designs. This research followed an embedded design where one data set, the student exemplars, is considered to be the primary data source, with the data collected from interviews and questionnaires considered as secondary data sources. Embedded designs may collect primary and secondary data simultaneously or sequentially; in this research sequential collection was used (Punch, 2009). Seeking a more holistic description can help to validate data collected from a variety of different sources.

A final quality consideration is the construction of the framework against which student work was assessed. This framework was a synthesis of a number of existing frameworks selected by the researcher. The selection of frameworks considered and the elements within these frameworks utilised in the synthesised framework was made by the researcher and as such may exhibit researcher bias.

5.6 Overview of ethical considerations

Research in mathematics education inevitably involves collecting information from and about people. Such an area of research will inevitably pose a number of ethical issues. Many professional bodies have developed their own ethical codes of conduct for research. Huberman and Miles (1994) identified 11 ethical issues in educational research. These issues arise at various stages of the research.

Worthiness of the project. Can this research make a valuable contribution to existing research conducted in this area? The research question must be able to withstand scrutiny and not simply be an indulgence of a researcher's preferred area of research.

Competence boundaries. Researchers, particularly those new to research, must question carefully their levels of competence and recognise that support or training may be required in some aspects of the research. The researcher must resolve to seek that support when necessary.

Informed consent. At all stages of the research process, participants have the right to be informed about any aspect of the research, including the right to withdraw at any stage. Participants should be aware of how their data will be collected, stored and used as well as knowing the motivation for the research.

Benefits, costs and reciprocity. Researchers need to be aware that although the research may advance their career in some way, there are often no obvious benefits for the participants, as their contribution is usually anonymous. Researchers need to be alert to possible benefits for participants such as giving them a chance to voice their concerns, the opportunity to reflect on practice or perhaps that results of the research may help participants to implement change.

Harm and risk. Researchers must always question whether there is any harm or risk, real or perceived, to participants in their research. Will participation in the research affect an individual's position at work or at home in any way? Will participation in the research be viewed negatively by an individual's colleagues or superiors? The level of harm or risk may vary from participant to participant, so researchers may need to answer these questions on a case by case basis.

Honesty and trust. Researchers need to establish a relationship with their participants that reflect honesty and trust. This relationship extends beyond the initial interview or survey to the objective reporting of results in a final document. Any misrepresentation of the data collected may adversely affect future research opportunities.

Privacy, confidentiality, and anonymity. Participants must be assured that their data will remain confidential and anonymous. Researchers must be aware that some details collected may unwittingly identify a participant. For example, referring to the gender of a participant from a particular school which has only one teacher of that gender at the school. Participants should be given precise details about exactly how their data will be used, analysed, published and stored. Such knowledge aims to reduce any privacy concerns of participants.

Intervention and advocacy. Sometimes during the process of research examples of harmful, illegal or wrongful behaviour may be observed. In such instances the researcher needs to decide whether or not to intervene and if they choose to intervene which side will they support. This may lead to some difficult ethical dilemmas for a researcher and if such examples are likely in their research they are advised to make decisions about possible courses of action early in the research process.

Research integrity and quality. Any researcher needs to question whether their methods bear scrutiny from their peers or from a wider audience. Researchers have a duty to faithfully record and report their research

Ownership of data and conclusions. Researchers may consider that the data and conclusions drawn from that data belong to them. However, interest groups have been known to block dissemination of reports if they are not happy with aspects of the findings. Such concerns may be mitigated by clear statements about ownership of data and future dissemination.

Use and misuse of results. A clear statement about possible uses of results from the outset of the research may help to avoid misuse in the future. A researcher might also need to decide what action they will take if they find their results have been deliberately misused or misinterpreted.

Heightened awareness of ethical issues at all stages of the research process may assist the researcher in either avoiding ethical dilemmas in the first place or being better prepared to deal with any dilemmas should they arise.

5.7 Ethical considerations in this research

The ethical considerations, using the criteria described by Huberman and Miles (1994) are now discussed with reference to this research project. It is important to note that approval of the University of Auckland Human Participants Ethics Committee (UAHPEC) was obtained for this research. The code of ethics used by this committee has been developed over a considerable period of time and has guided numerous research projects over a wide range of disciplines.

One of the early catalysts behind this research was NZQA who registered an interest in any research projects about their recently introduced Achievement Standards. I considered my competence to be of an acceptable level since I had completed a post-graduate paper on research methods as part of a post graduate diploma of science. In the planning stages of the research a number of other ethical issues emerged. For example, initially I had hoped to use student exemplars from the selected secondary schools for both Achievement Standards. However, as this is an Achievement Standard that is completed in the last year of school, gaining the consent of students who had already left school was eventually deemed to be too time-consuming, if indeed it was possible. Using these exemplars which schools regularly retain for several years after the assessment without the consent of the authors was considered ethically unsound and so was not pursued.

The initial approaches to each school was to the Principal and only once their consent was received were the Year 13 statistics teachers approached at that school and invited to complete a questionnaire and possibly be interviewed. It was made clear to potential participants that their involvement in the research was entirely voluntary and would not result in any negative repercussions for them professionally. Any element of coercion of participants is considered ethically unsound. Participants were not offered any remuneration for their co-operation but I hope they found the opportunity to reflect on their teaching practices beneficial and the research results in general may assist in their teaching of time series in the future. Principals and teacher participants were all sent an information sheet and asked to sign a consent form (see Appendices 3 and 4). All participants were offered the right to withdraw from the research at any time without specifying a reason.

Data collected during this research was only seen by me and my supervisor. The only exception to this was the interview data which was transcribed by a qualified transcriber who signed a confidentiality form. Anonymity of all participants was also assured. Student exemplars used were freely available on the internet and were anonymous. Participants were not considered to be at any risk during this research.

At the conclusion of the project, there are ethical considerations regarding ownership of data and dissemination of reports and conclusions. These are all covered by guidelines issued by the University of Auckland Human Subjects Ethics committee. Data will be stored securely in the Department of Mathematics for a period of six years, after which it will be destroyed. All participants were offered the option of receiving a copy of the thesis on completion.

5.8 Summary

This chapter has briefly discussed a variety of research methods commonly used in education research and then described more fully which of these methods were utilised in this particular research project. In selecting and implementing a research method, questions of validity and quality have also been considered and careful attention has been paid to ethical issues.

- Both quantitative and qualitative data were collected. This mixed methods approach was adopted in order to gain a more holistic perspective of the teaching of time series analysis in New Zealand secondary schools rather than solely an analysis of student responses to assessments. This type of research is known as an embedded triangulation design.
- Exemplars of student work were accessed from freely available material available on the internet. These exemplars were uploaded by NZQA prior to the introduction of each Achievement Standard into the New Zealand curriculum. Additional exemplars were circulated at professional development workshops. As such the consent of the authors of this work did not need to be sought. These exemplars were assessed against a synthesised framework (see Chapter Four).
- Participants in this research were fully informed about their commitment and all voluntarily agreed to engage in the project. All data collected from the participants will be securely kept for a maximum of six years after which it will be destroyed. Participants were made aware that data collected will be published in a thesis and may also be mentioned in future presentations or papers.
- Of an initial nine schools selected to take part in the research, one declined and three failed to return any questionnaires despite numerous reminders. This left five schools at which the lead Year 13 statistics teacher was invited to participate in an interview of not more than an hour in length. Teachers at all five of these schools agreed to be interviewed.
- Ethical concerns are paramount in any research involving human participants. In this research schools were assured that neither they nor their staff would be identifiable. No element of coercion was used to encourage participation. Confidentiality regarding all data collected was also assured.

CHAPTER SIX

Results

6.1 Introduction

The results presented in this chapter derive from three sources, exemplars of student work, teacher questionnaires and teacher interviews. Student exemplars of work completed prior to the introduction of AS 3.1 and AS 3.8 were examined and evaluated against the synthesised framework discussed in Chapter Four. The teacher questionnaires provided some background information on how teachers and their schools have managed the change from AS 3.1 to AS 3.8. Topics included resources, management of assessment, changes in use of technology and professional development as well as any changes in student responses. The teacher interviews sought to gain further insight into these differences and to provide a more holistic view of any shift in the teaching of time series in the school curriculum.

6.2 Student exemplars

A total of thirty four exemplars of student work, generating over six hundred cases of levels of reasoning, were compared against the framework described in Chapter Four. The exemplars used were obtained from the internet; freely available and uploaded by NZQA prior to the introduction of AS 3.1 and AS 3.8. Some further AS 3.1 exemplars, which had been used in moderation workshops, were also supplied by NZQA. Some of the exemplars were classified by achievement level, others were not and some were only part-assessment responses. The part-assessment responses are included in the overall analysis but not the analysis by achievement level.

6.2.1 Revision of framework

For each student exemplar every calculation, comment and equation was categorised according to the framework. If a comment or calculation contained something incorrect it was still included in the overall analysis of level of reasoning. In the process of coding, the original framework was extended; seven new codes were added to existing levels of reasoning in order to code unanticipated responses. The six main levels of reasoning (vertical, horizontal, procedural, extended procedural, interpretive and interrogative) remain but additional sub-categories were added in levels 2, 3, 4 and 5 (Table 6). Although the overall levels of reasoning are designed to be hierarchical in nature, the sub-category codes are not. These additional codes do not reflect higher levels of reasoning at their respective levels but rather different ways of demonstrating that level of reasoning.

Table 6 Revised framework - additional sub-category codes added

Level	Level of Reasoning	Code	Description
2	Horizontal Reasoning	2.4	Comment on patterns in variation
3	Procedural Reasoning	3.8	Calculation of seasonally adjusted data
3	Procedural Reasoning	3.9	Discuss model goodness of fit
3	Procedural Reasoning	3.10	Graphical production of residuals
4	Extended Procedural Reasoning	4.4	Discuss validity of model
5	Interpretive Reasoning	5.4	Identification of interest groups
5	Interpretive Reasoning	5.5	Interpretation of residuals

No examples of horizontal reasoning were found in the exemplars but this level of reasoning is retained in the framework for the sake of completeness. If the framework is to be used to assist in developing the teaching of time series then this level is an important first step. Table 7 displays the revised framework including the additional sub-category codes.

Table 7 Revised coding framework

Level	Level of reasoning	Characteristics	Codes	Description in time series	Theoretical frameworks
1	Vertical reasoning	Ability to read and understand scale, units and basic graph construction. Context may be present but not critical to task.	1	Read individual data values from a graph.	Konold - data as a pointer, data as a case value Curcio - reading the graph SOLO- concrete-symbolic, unistructural RME - context free/camouflage, low level reasoning, simple complexity
2	Horizontal reasoning	Looking across the data not just individual data values. Context may be present but not critical to performance of tasks.	2.1	Read across the data to assess overall trend direction	Konold - data as an aggregate
			2.2	Identification of when peaks and troughs occur	Curcio - reading within the graph SOLO- concrete-symbolic/formal, unistructural/multi-structural
			2.3	Identification of any unusual values	RME - context free/camouflage, low level reasoning, simple complexity
			2.4	*Comment on patterns in variation*	Wild & Pfannkuch – consideration of variation
3	Procedural reasoning	Use of mathematical procedures to calculate properties of data set or to examine features of data set. Context may be provided but not critical to performance of procedure.	3.1	Calculation of smoothed data	RME - context free/camouflage, low/middle level reasoning, middle complexity
			3.2	Graphical production of smoothed data	SOLO- formal, unistructural
			3.3	Calculation of seasonal effects	SOLO- formal, unistructural
			3.4	Graphical production of seasonal effects	SOLO- formal, unistructural
			3.5	Find equation of line of best fit	SOLO- formal, unistructural
			3.6	Trend estimation using line of best fit	RME - context free/camouflage, low/middle level reasoning, middle complexity SOLO – formal, unistructural
			3.7	Trend estimation using smoothed data	SOLO- formal, unistructural RME - context free/camouflage, low/middle level reasoning, middle complexity
			3.8	*Calculation of seasonally adjusted data*	SOLO- formal, unistructural RME - context free/camouflage, low/middle level reasoning, middle complexity
			3.9	*Discuss model goodness of fit*	SOLO- formal, multi-structural
			3.10	*Graphical production of residuals*	SOLO- formal, unistructural
4	Extended procedural reasoning	Using a mathematical model to predict beyond the current domain of the data set or to interpolate within the give domain. Context may be provided but not critical to performance of procedure.	4.1	Use mathematical model to make predictions beyond current data set	Curcio - Reading beyond the graph. SOLO - formal, multi-structural
			4.2	Discuss validity of predictions	RME - context free/camouflage, middle level reasoning, middle complexity
			4.3	Use mathematical model to find missing values within domain of current data set	RME - context free/camouflage, low/middle level reasoning, middle complexity
			4.4	*Discussion of validity of model*	RME - context free/camouflage, middle level reasoning, middle complexity

5	Interpretive reasoning	Ability to interpret implications and characteristics of a mathematical model. Transnumeration - examination of different representations of the data to aid understanding. Context must be present and any interpretation provided must refer to it. Consideration of variation.	5.1	Describe contribution of various components of time series and attempt possible interpretation in context	Shaughnessy - reading behind the graph
			5.2	Identification of variation in trend and data over time and provision of possible interpretation in context	RME - authentic context, middle/high level reasoning, middle/high complexity
			5.3	Some consideration of other factors to assist in seeking for explanations of variation in time series	Wild & Pfannkuch - investigative cycle (PPDAC), interrogative cycle, types of thinking including transnumeration, consideration of variation and reasoning with a statistical model
			5.4	*Identification of interest groups*	RME - authentic context, middle/high level reasoning, middle/high complexity Shaughnessy - reading behind the graph
			5.5	*Interpretation of residuals*	RME - authentic context, middle/high level reasoning, middle/high complexity Shaughnessy - reading behind the graph
6	Interrogative reasoning	Seeks information to assist in interpretation or that may confirm or refute own findings. Ability to critique own analysis and to raise further questions. Have a healthy dose of scepticism. Integration of contextual and statistical knowledge essential.	6.1	Examination of other research or findings to confirm or refute own findings.	Wild & Pfannkuch - investigative cycle (PPDAC), interrogative cycle, types of thinking including transnumeration, consideration of variation and reasoning with a statistical model
			6.2	Detailed consideration of other factors to assist in seeking for explanations of variation in time series	Wild & Pfannkuch - dispositions including scepticism, curiosity and perseverance
			6.3	Acknowledgement that their analysis is not exhaustive and that other interpretations may be valid.	SOLO - formal/post formal, relational/ extended abstract
			6.4	Integration of statistical and contextual knowledge	RME – authentic context, high level reasoning, high complexity

Additional codes in italics.

6.2.2 Analysis of literary statements

Any cursory analysis of the student exemplars reveals a difference in styles between responses to the two standards. Exemplars for AS 3.8 in the main are submitted as a short report whilst those for AS 3.1 have a greater emphasis on presentation of calculations. The report style of AS 3.8 exemplars meant that there were a large number of different types of literary statements made in responses that could not be coded using existing codes in the original or revised framework. Over one hundred (24%) such statements were found in exemplars from AS 3.8 from a total of 484 cases of reasoning levels. Only eight (4.5%) similar types of literary statements were found in exemplars from AS 3.1 out of a total of 176 cases (Figure 6). Exemplars from AS 3.8 for the most part followed the PPDAC cycle, a structure which is familiar to most NZ secondary school students (Wild & Pfannkuch, 1999). A typical AS 3.8 report would start with an opening statement about the problem and why the investigation was interesting or worthwhile. A description of the important features of the data followed, illustrated by graphs. The analysis section would include further graphs, calculations and details of predictions. This section might also examine other related research findings and the report would end with some conclusions. Exemplars from AS 3.1 from the lower achievement levels, Not Achieved and Achieved, might only include one sentence explaining the long term trend and one sentence explaining the prediction in context. At higher levels, Merit and Excellence, a short paragraph describing the trend or features was often included. At Excellence level students were expected to include commentary on particular aspects of the time series which they selected from a given list. The options included: relevance and usefulness of forecast; appropriateness of the model; improvements to the model; limitations of the analysis; seasonally adjusted data; and comparison with related time series data.

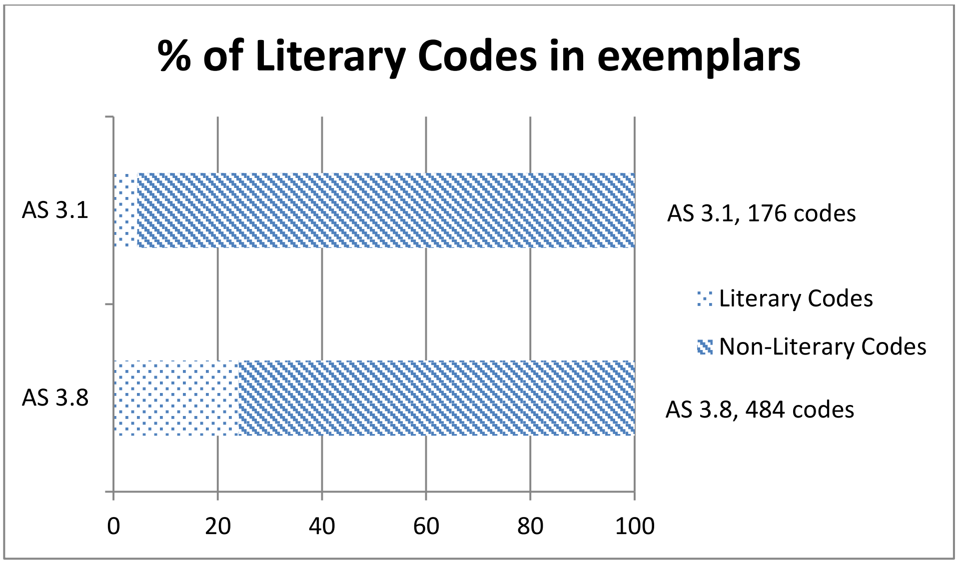


Figure 6 Percentage of literary codes found in AS 3.1 and AS 3.8 exemplars

Some student exemplars from AS 3.1 contained no literary statements at all just a series of calculations. Given this disparity in frequency of literary statements made in the two types of exemplars, these new types of literary statements were regarded as a separate but new component. Four new codes were created for such literary statements but these codes were *not included* in the revised framework and were *not included* in any analysis of results other than those reported in this section. Inclusion of these codes of literary statements in the overall analysis would skew comparisons of the original levels of reasoning.

Four new codes were established in order to categorise these literary statements found predominantly in student exemplars produced for AS 3.8. Descriptions and examples of these codes are provided in Table 8.

Table 8 Literary statement codes

Code	Description	Examples from student exemplars
T1	Scene setting comment	*The Arctic sea ice is the result of the sea water in the Arctic Ocean freezing when its temperature drops below -2℃ (freezing point of sea water).*
T2	Purpose statement	*I am going to investigate the mean area of sea ice in the Arctic to see if this decrease over time is actually happening and how fast it is occurring*
T3	Definition or description of data	*The mean area of the Arctic[sea ice] was calculated from daily satellite images from January 1990 to March 2011*
T4	Sentence in words following a calculation	*I predict that the amount of Arctic sea ice for September 2011 will be 14.72 million square km.*

When considering the revised framework some codes for literary statements are included, for example, code 2.4 ‘comment on patterns in variation’ or code 4.2 ‘discuss validity of the model’. An analysis was conducted which considers the differences for *all* literary codes found in the student exemplars and results are presented in Table 9.

Table 9 Comparison of frequency and percentage of literary codes in student exemplars from AS 3.1 and AS 3.8

All Literary Codes	AS 3.1 (Frequency)	AS 3.8 (Frequency)
2.2, 2.3, 2.4	7	44
3.6,3.7,3.9	15	13
4.2,4.4	34	62
5.1 to 5.5	17	79
6.1 to 6.4	0	22
T1 to T4	8	116
Total literary codes	**81**	**336**
Total all codes	176	484
% Literary codes	*46%*	*69%*

Thus even when all codes pertaining to literary statements are included the same shift towards more literary statements in students exemplars from AS 3.8 persists, with 69% of all codes in AS 3.8 exemplars representing some type of literary statement in contrast to only 46% of all codes in AS 3.1 exemplars.

The analysis in the following sections focuses on the original levels of reasoning described in full in Chapter Four including the additional sub-categories introduced in the revised framework (Table 7). Literary codes, T1 to T4, are shown in the coding process but are *excluded* from the analysis.

6.2.3 Coding of student exemplars

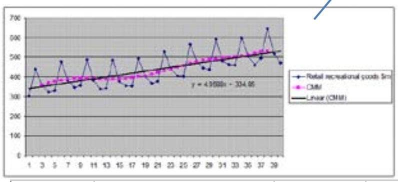
In this section two examples of coding of student exemplars are demonstrated, with an explanation of how the framework categories in Table 7 and Table 8 were applied. Both examples are from Merit achievement levels, one from AS 3.1 and one from AS 3.8. I have chosen two examples that use the same software, i.e. EXCEL.

Example of coding a Merit level student response under AS 3.1 using EXCEL

Student Response

3.1 – Calculation of smoothed data

3.3 – Calculation of seasonal effects



The line on the graph shows the long term trend. From this straight line it implies that the sales of recreational goods are increasing by about $4.96 million each quarter.

First sentence coded T3 (description). Second sentence coded 3.6 – trend estimation using line of best fit.

Quarter	Retail Recreational Goods $m	CMM	SE
Sep-95	303		
Dec-95	440		
Mar-96	358	359.75	-1.75
Jun-96	324	368.125	-44.125
Sep-96	331	376	-45
Dec-96	479	381.625	97.375
Mar-97	382	387.5	-5.5
Jun-97	345	391.75	-46.75
Sep-97	357	392.875	-35.875
Dec-97	487	392.125	94.875
Mar-98	383	389.5	-6.5
Jun-98	338	387.5	-49.5
Sep-98	343	386.5	-43.5
Dec-98	485	387.75	97.25
Mar-99	377	391	-14
Jun-99	354	393.25	-39.25
Sep-99	353	397.5	-44.5
Dec-99	493	402.25	90.75
Mar-00	403	406.625	-3.625
Jun-00	366	414	-48
Sep-00	376	422.875	-46.875
Dec-00	529	432.125	96.875
Mar-01	438	440.5	-2.5
Jun-01	405	448.5	-43.5
Sep-01	404	458.25	-54.25
Dec-01	565	468.375	96.625
Mar-02	480	477.375	2.625
Jun-02	444		
Sep-02	437		
Dec-02	592		
Mar-03	480		
Jun-03	463		
Sep-03	459		
Dec-03	598	506	92
Mar-04	506	510.25	-4.25
Jun-04	459	521.125	-62.125
Sep-04	497	529	-32
Dec-04	647	532.25	114.75
Mar-05	520		
Jun-05	471		

Seasonal Effects	
March	-5.69444
June	-45.5972
Sept	-44.1667
Dec	97.98611

Calculations coded as 4.1 – using model to predict beyond current data set. Sentence coded as T4 – words following a calculation.

Predictions

December 06

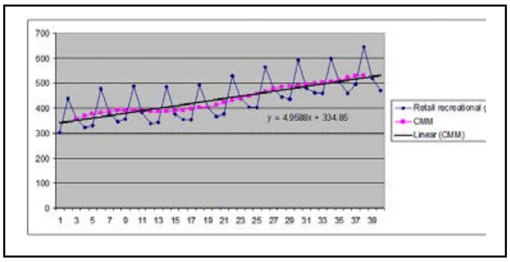
4.9588 × 46 + 334.85 + 97.98611 = 660.9409

The estimated value of the recreational sales in December 2006 will be $661 million.

June 07

4.9588 × 48 + 334.85 - 45.5972 = 527.2752

The estimated value of the recreational sales in June 2007 will be $527 million



3.5 - Equation of line of best fit found

Example of coding a Merit level student response under AS 3.8 using EXCEL

First five sentences coded as T1 – scene setting. Final sentence coded as T2 – purpose statement.

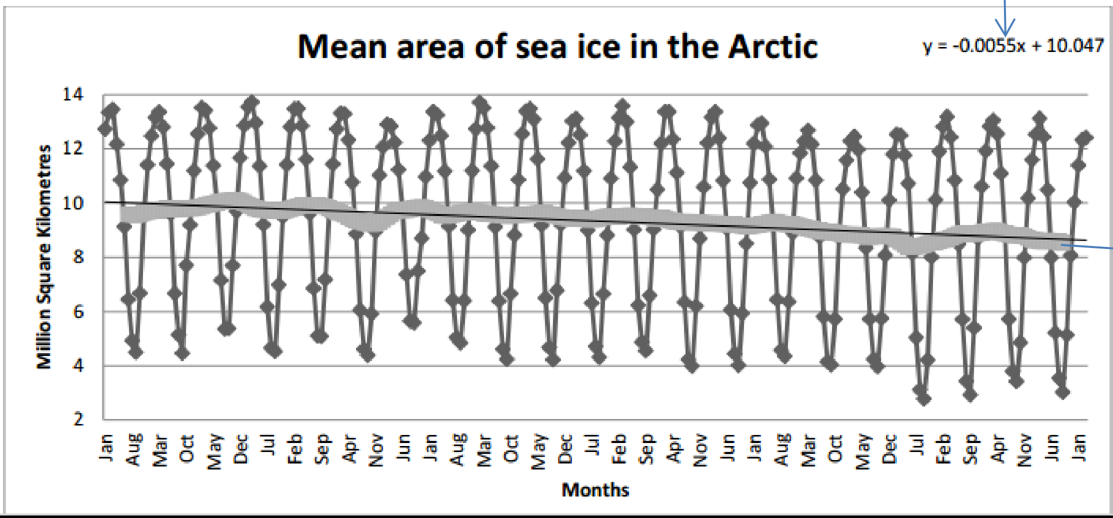
First sentence coded as 2.1 – reading across the data. Sentences 2 and 3 coded as 5.2 – possible interpretation of trends offered. Last sentence coded as T3 – data description.

It is well known that global warming is occurring on our planet. Global warming is the rise in the average temperature of the Earth's atmosphere and oceans. These temperature rises could then cause devastating effects to the wildlife and their habitats on Earth. There have been reports that the area of sea ice in the Arctic has been decreasing because of global warming and so could mean that the habitats of the polar bears that live in the Arctic could be diminished fairly soon. I am going to investigate the mean area of sea ice in the Arctic to see if this decrease over time is actually happening and how fast it is occurring.

Features of the raw data

The raw data shows the mean area of sea ice in the Arctic is slightly decreasing due to global warming. Polar sea ice helps reflect solar heat back into space, making the air colder, while reducing heat loss from the water below the ice and controlling currents. However CO[2] emissions create a carbonated cloud which blocks the sun rays from exiting the earth causing global warming. The mean area of the Arctic was calculated from daily satellite images from January 1990 to March 2011.

3.5 – Calculation of line of best fit



3.1 – Calculation of smoothed data

There is a Seasonal pattern happening to the mean area of sea ice in the Arctic as the peaks occur during the Winter months (Jan-Apr) when the temperatures are lower whereas the troughs occur during the Summer months (Jul-Oct) when the temperatures are higher. This is because sea ice freezes during the cold winter months and it shrinks during the summer months as the sea ice melts.
There is a pretty stable variability from the mean area of sea ice data in the arctic as it appears to only be slightly decreasing in value over time.
The gradient of the trendline suggests that the mean area of sea ice in the arctic is slightly decreasing at a rate of 66,000 square kilometres per year on average.
There are no known spikes shown in the graph.

First sentence coded 2.2 – identification of peaks/troughs. Second sentence coded 5.2 – interpretation of variation in context. Third sentence coded 2.4 – comments on variation. Fourth sentence coded 3.6 – trend estimation using line of best fit. Last sentence coded 2.3 – identification of peaks/troughs.

Forecasts

Each prediction calculation coded as 4.1 – using model to predict beyond range of current data. Each following sentence coded as T4 – words following a calculation.

Sept 2011, x = 261

Y = -0.0055 x 261 + 10.047 -5.10 = 3.5115

My prediction for September 2011 suggests that the estimate for the mean area of sea ice in the Arctic is 3.51 million square kilometres.

March 2012, x = 267

Y = -0.0055 x 267 + 10.047 + 3.84 = 12.4185

My prediction for Mar-12 suggests that the estimate for the mean area of sea ice in the Arctic is 12.42 million square kilometres.

Since the trend line appears to fit the smoothed data reasonably well with very little deviation away from the CMM line (apart from April-July 1995 and April-July 2007) so my predictions should be fairly are reliable

The data from satellite observations of the mean area of sea ice in the Artic is covered for 20 years of monthly data and used to make predictions for the further years and therefore making my predictions more accurate.

"Since the trend line…." And "The data from satellite….." coded as 4.2 – discussion of prediction validity.

The data for mean area of sea ice in the Arctic only shows the surface mean area of the Arctic. However this is not enough to tell if global warming is affecting it as you may need to know the depth of the Arctic sea ice as well as the temperature of the water below the ice. Also there should be data for CO^2 emissions as it controls how harsh the summer months are getting when the sun rays gets trapped inside the earth, and therefore accelerating the process when the polar ice caps melt.

First sentence coded as T3 – data description. Second & third sentences coded as 5.3- some consideration of other factors but not related to features observed in the time series.

All exemplars were coded in a similar fashion in order to assess the level of reasoning demonstrated in the student responses. Table 10 summarises these results showing the percentage of response items that were categorised at each level of reasoning.

Table 10 Overall level of reasoning shown in student exemplars produced for AS 3.1 & AS 3.8, excluding literary codes

Level of Reasoning	% at each level, AS 3.1 exemplars	% at each level, AS 3.8 exemplars	Difference (3.1 – 3.8)
1- Vertical	0	0	0
2- Horizontal	9.5	26	-16.5
3- Procedural	47	21	26
4- Extended Procedural	33.5	26.4	7.1
5- Interpretive	10	20.6	-10.6
6- Interrogative	0	6	-6
Total	100	100	

The vast majority of student responses to AS 3.1 were at level 4 or below (90%). This compares with 73% under AS 3.8. Only 10% of student response types under AS 3.1 reached a level of 5 and none reached level 6. Under AS 3.8, 20.6% of student responses reached level 5 and 6% reached level 6. So although over half the responses are at level 4 or below, the procedural levels, AS 3.8 shows a distinct shift in student responses towards the higher levels of reasoning. It is also interesting to note that under AS 3.8 more emphasis has been placed on level 2, horizontal reasoning, which indicates that more teaching effort and time is going in to trend description.

6.2.4 Comparison of mean reasoning level by level of achievement

If a null hypothesis is considered of no difference in the mean level of reasoning demonstrated between the two standards, then the difference in means observed at each achievement level should be zero. Any departure from zero provides evidence against the null hypothesis, which would support the conjecture that a difference in the level of reasoning between the two standards does exist.

The codes from this data are not independent since each exemplar yields a number of codes from one individual. Standard statistical tests such as the t-test also require an assumption of normality which again is not supported with this data. However, it is still interesting to compare the mean level of reasoning by level of achievement in each standard. Table 11 shows a mean level of reasoning for each achievement level and for each standard. Mean levels of reasoning, as might be expected, tend to increase in both standards with the level of achievement. The exception being that the mean level of reasoning in AS 3.1 exemplars is virtually identical in Merit and Excellence exemplars. However, the differentiation between levels of reasoning by achievement level is greater in the new standard. Overall in AS 3.8, the mean level of reasoning is greater at every achievement level, with the biggest increase (17.4%) in level of reasoning found at the Excellence achievement level. Achievement and Merit levels showed increases in the mean level of reasoning of between 5 and 10 per cent respectively.

Table 11 Comparison of mean levels of reasoning in student exemplars between AS 3.1 and AS 3.8

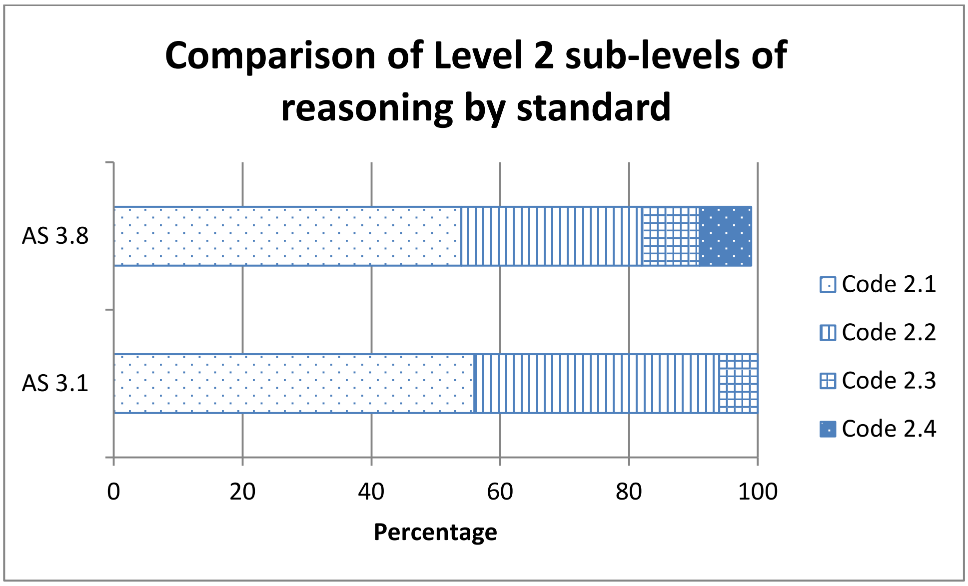
		Level of Reasoning					
	No. Exemplars	2	3	4	5	6	Mean
AS 3.1							
Achieved	4	1	15	0	0	0	2.94
Merit	6	9	34	30	8	0	3.41
Excellence	5	3	25	16	3	0	3.40
AS 3.9							
Achieved	6	35	28	27	16	0	3.23
Merit	6	29	20	38	34	0	3.64
Excellence	3	28	18	26	29	22	3.99

6.2.5 Analysis of sub-code levels of reasoning

However, Table 11 considers the broad levels of reasoning only and does not examine any shift that may or may not have occurred within levels of reasoning. The categories within each framework level were not intended to be hierarchical, yet the labelling if used to compare these two distributions would assume a hierarchy is present. For example, that code 3.9 represents a higher level of reasoning than code 3.1. In some cases this was the case but in others the levels might be regarded as comparable but different. For example, code 3.1 represents calculation of smoothed data and code 3.2 represents graphical production of smoothed data. Both result in the same outcome and neither can be argued as demanding a higher level of reasoning.

Figure 7 illustrates the level of reasoning found in each Achievement Standard at level 2, horizontal reasoning and at level 3, procedural reasoning. There were 16 cases of level 2 reasoning in AS 3.1 exemplars compared to 96 cases in AS 3.8. Proportions of code 2.1 – reading across the data, were similar in both standards. The proportion of code 2.2 was slightly higher in AS 3.1, 38%, compared to only 28% in AS 3.8. No examples of code 2.4 – comment on patterns of variation, were found in AS 3.1 exemplars but 8% of AS 3.8 level 2 codes were in this category.

A total of 79 cases of level 3, procedural reasoning, were found in AS 3.1 exemplars and 77 cases found in AS 3.8 exemplars, but the proportions of each type of sub level of code varied. For example, almost 60% of the AS 3.1 level 3 codes were either 3.1 – calculation of smoothed data or 3.5 – find equation of line of best fit. Whereas in AS 3.8 exemplars there was a more even spread across codes 3.1 to 3.6 with around 15% in each category.



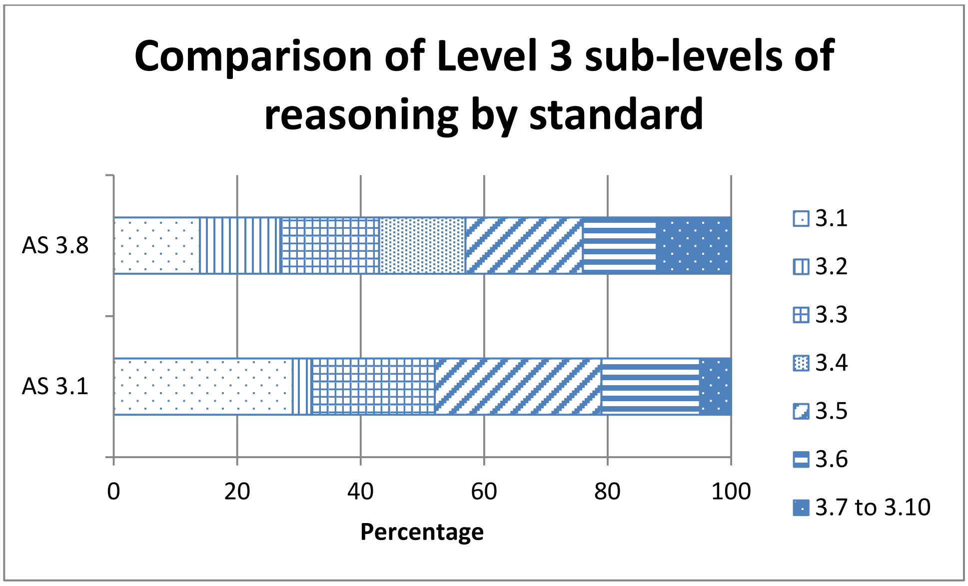


Figure 7 Comparison of level 2 and 3 sub-level codes of reasoning by standard

At level 4, extended procedural reasoning, 93% of codes in AS 3.1 exemplars were either code 4.1, use mathematical model to make predictions, or code 4.2, discuss validity of predictions. In AS 3.8 exemplars, these codes also dominated (79%) and a further 19% of cases were coded, 4.4, discuss validity of model.

A greater diversity in the sub-levels of reasoning appears to have been activated in AS 3.8 when compared with AS 3.2 in the categories of horizontal and procedural reasoning. This greater diversity is not found with sub- levels of reasoning at level 4, extended procedural reasoning.

6.2.6 Levels of reasoning by achievement level

The levels of reasoning demonstrated in the student exemplars can also be investigated by each NCEA achievement level. Five student exemplars which were not categorised by achievement level are excluded from the following analysis. Under both standards there were three different levels of achievement – Excellence, Merit and Achieved. Figure 8 compares levels of reasoning for the Excellence achievement level.

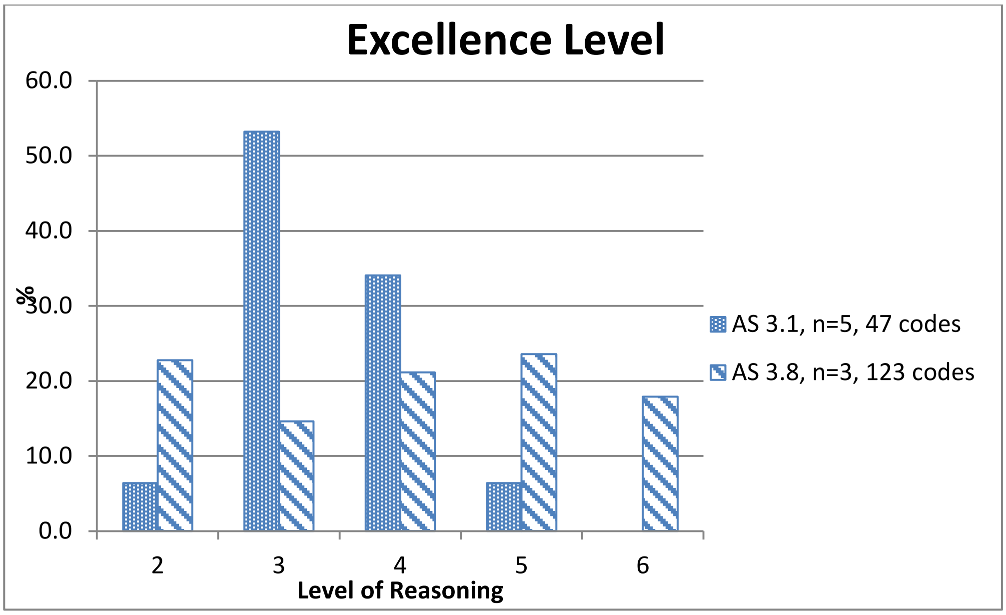


Figure 8 Percentage of Excellence student exemplar responses by level of reasoning, AS 3.1 and AS 3.8

The level of reasoning demonstrated under AS 3.1 at Excellence level is predominantly at level 3, procedural reasoning, just over a third of responses reached the extended procedural level, level 4. Only 6.4% reached level 5, the interpretive level and none reached the interrogative level. Unlike responses under AS 3.8, which show a wider range of levels covered with over 40% of responses at the highest two levels of reasoning.

Student exemplars that were graded at the Merit level (Figure 9) largely exhibited level 3 or 4 procedural or extended procedural reasoning. In time series such a level of reasoning requires students to be able to follow procedures for calculation of or graphical production of smoothed values and predictions. Such procedures may be carried out without reference to the context of the data and without any attempt to interpret what they have produced. Often students rote learned these procedures; Teacher 4 described such skills as "a knowledge that is gone as soon as the test is over". AS 3.8 student responses at Merit level did display slightly higher levels of reasoning than AS 3.1, with 60% of responses at level 4 or 5. Under AS 3.1, 47% of responses reflected a level of reasoning at level 4 or 5. It is interesting to note that the percentage of level 5 codes in AS 3.1 is higher at Merit level than it is at Excellence level. If we examine individual Excellence and

Merit scripts, this is explained by one particular exemplar which contained five level 5 codes and was a borderline Excellence script.

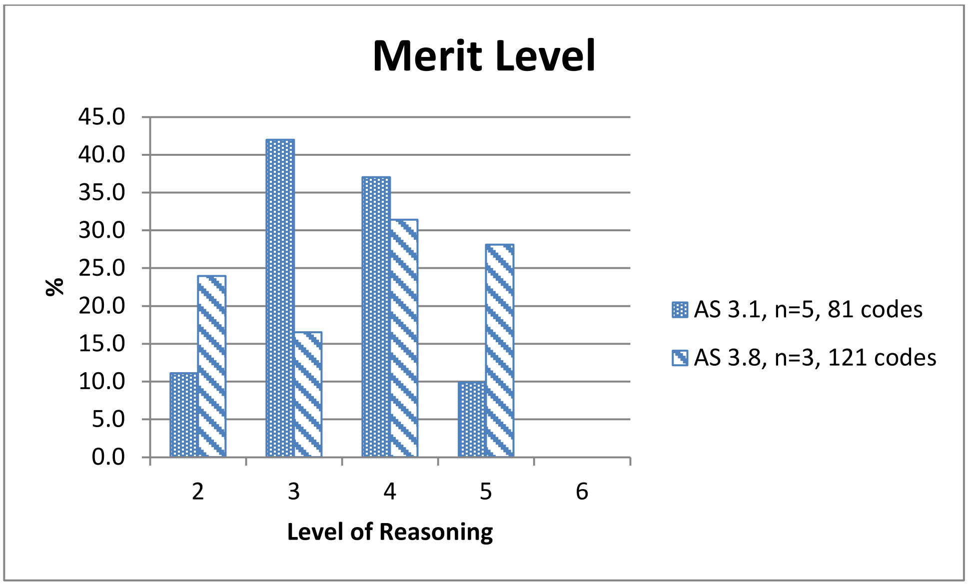


Figure 9 Percentage of Merit student exemplar responses by level of reasoning, AS 3.1 and AS 3.8

The majority of responses at the Achieved level (Figure 10) under AS 3.1 were almost exclusively at the procedural level of reasoning, level 3. It is interesting to note that under AS 3.8 even at the Achieved level there was some evidence of reasoning at higher levels, extended procedural and interpretive reasoning, whereas under AS 3.1 there was none.

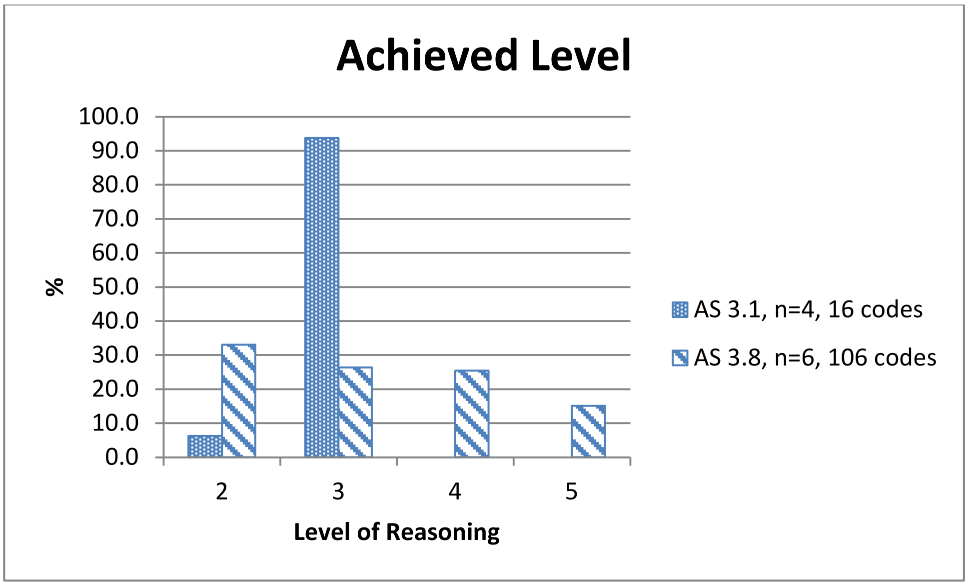
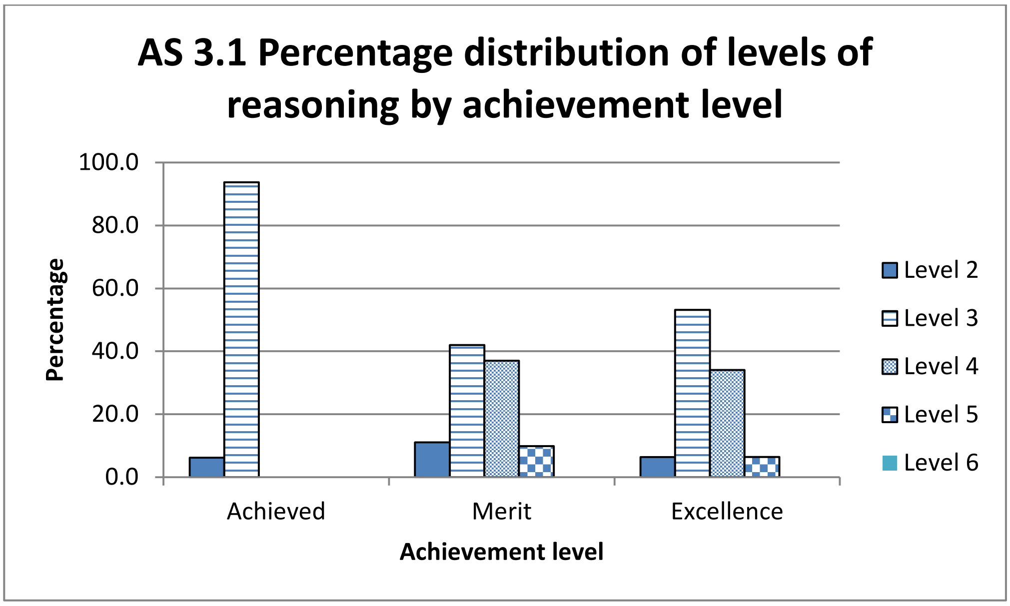


Figure 10 Percentage of Achieved student exemplar responses by level of reasoning, AS 3.1 and AS 3.8

Figure 11 combines Figures 8, 9 and 10 and demonstrates the increased incidence of higher levels of reasoning in AS 3.8 exemplars compared to AS 3.1 exemplars. A broader range of levels of reasoning is observed at all achievement levels, but this is particularly noticeable at the Excellence level.



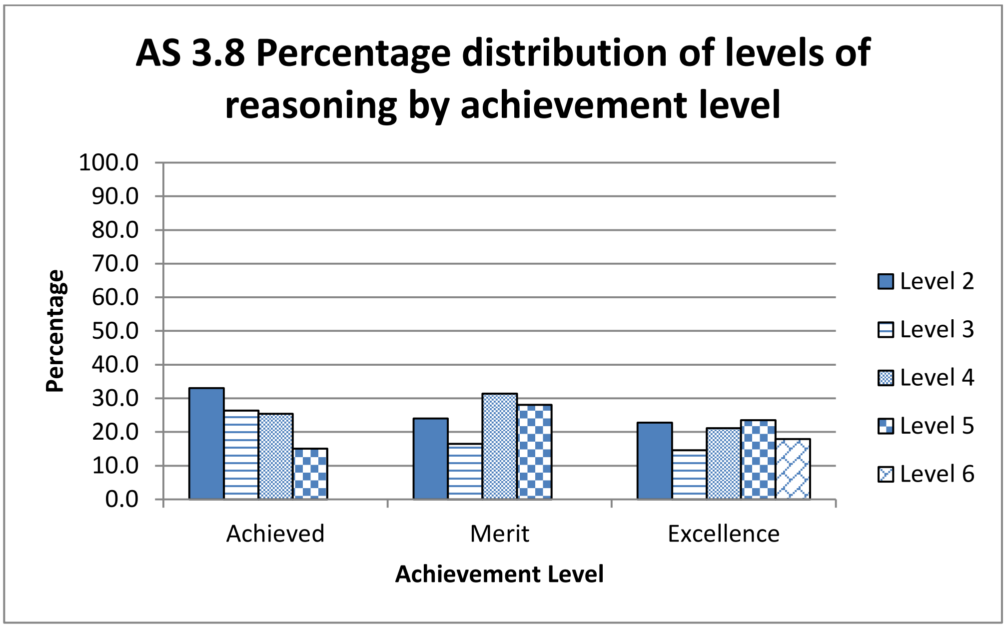


Figure 11 Comparison of reasoning levels across achievement levels and by standard

6.3 Teacher questionnaires

A total of eighteen completed questionnaires were received from teachers at six different secondary schools. The teachers were from a mixture of State (14) schools and private (4) schools, all with a decile ranking of 7 or higher. Some lower decile schools were invited to take part but declined the invitation. There were no rural or all boys' schools. Twelve teachers were from co-educational schools and six were from girls only schools. Six teachers reported a high proportion of English as second language speakers at their school. Teachers described their

Mathematics Department as fairly traditional (6), a mixture of traditional and progressive (6) or fairly/very progressive (6).

The experience of the teachers who completed the questionnaires was varied as shown in Table 12.

Table 12 Experience of teachers who completed questionnaire

	HOD	Teacher in charge (TIC) Y13 Stats	TIC of another level	Y13 stats teacher
Position held	5	4	4	5
Teaching qualification	**From NZ**	**From Overseas**		
Frequency	16	2		
Years teaching Y13 Stats	**Less than 5 years**	**6 to 10 years**	**Over 10 years**	
Frequency	2	6	10	
Years teaching time series	**Less than 5 years**	**6 to 10 years**	**Over 10 years**	
Frequency	3	9	6	
Time series study level of study	**Post-graduate**	**Undergraduate**	**Neither**	
Frequency	2	4	12	

Predominantly the views expressed in this questionnaire were from teachers with considerable teaching experience who hold positions of responsibility within their departments. Most had taught Year 13 statistics for over 10 years and time series for between 6 and 10 years. All teachers had experience of teaching both AS 3.1 and AS 3.8, but only six teachers had formally studied time series analysis as part of their own qualifications, with two of these having tackled the topic at post graduate level. This result was anticipated, thus the next section of the questionnaire sought to discover how teachers had acquired the time series knowledge required to teach, in particular, the new standard, AS 3.8. The most popular methods for teachers were to attend workshops at Statistics Days (17), or to utilise colleagues' knowledge (17). Some used on-line resources such as Census@School (13) or undertook some self-directed study of their own (13). Eleven teachers reported attending some form of professional development run by their school. Teachers were asked to rank how beneficial the forms of professional development they had attended were, with a rank of 1 being the most useful (Table 13).

Table 13 Types of professional development experienced for time series

Professional development source	Average rank score, rank 1 = most useful	Overall rank
Statistics days	1.875	1
Local maths association	2.28	2
Own school	2.375	3
Colleagues	3	4
Census@School	3.5	5
Other online resources*	3.6	6
NZQA workshop	3.7	7
Privately run workshop	4	8
Self-study	4	9
Conference workshop	4.75	10
Text books	6.8	11

*Included www.mahobe.co.nz, Nayland maths, Statistics NZ, Statistics Learning Centre, UoA *iNZight* online courses, Dr Nic, Priscilla Allan, J. Wills

Thus Statistics Days, professional development events at the end of each year, not only attracted the largest numbers they also rated highly in terms of efficacy. Seven teachers attended professional development sessions run by their local maths associations and these too were rated highly. Self-directed study, although one of the more frequent sources of professional development, was ranked as not particularly effective by teachers who had utilised this source.

Most teachers experienced some form of professional development prior to the introduction of AS 3.8, but as the responses to the following set of questions reveals, (Table 14) some did and some did not feel that this adequately prepared them for the teaching involved.

Table 14 Teacher preparation for the new AS 3.8 standard[1]

Question	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
Did you feel competent to teach AS 3.8?	4	11	1	1	0
Did you feel well-prepared to teach AS 3.8?	1	12	1	3	0
Did you feel confident teaching students how to structure and write reports?	0	13	0	3	1

1. One participant did not respond to these three questions.

The majority of these experienced teachers agreed or strongly agreed that their professional development had prepared them for teaching the new standard, although they were slightly less confident about literary aspects. Four teachers felt less well prepared despite all of these teachers having experienced at least two or more sources of professional development and all having attended Statistics Day workshops ranked by other teachers as the most beneficial.

Teacher 12 felt

> I really needed a solid year of background learning myself before I was let loose with a Year 13 class. MOE budgets never allow for this. PD [professional development] is always too fast, too little and too superficial – even if you do have time to go to what is offered.

Teachers were also asked to describe the workload of implementing AS 3.8; this ranged from substantial (9), somewhat substantial (7) and two were not sure. Comments were also sought on whether any obstacles were present that prevented them teaching time series as they wished. Responses to this were varied with six teachers citing the lack of availability of interesting time series data sets. Four teachers mentioned that as this standard is often the first internal of the year it has to compete with other start of the year events which means it is often rushed. Only two teachers mentioned availability of computers as an obstacle. Three teachers cited *iNZight* as an obstacle; one mentioned poor graphics, another felt that automatic production of graphics did not help students to understand features of time series and a third felt the calculations required for AS 3.1 helped students to interpret times series better. One teacher lamented the lack of adequate published resources for this topic, although it is acknowledged that publishers had to prepare much of this material before AS 3.8 had been finalised.

Finally in this section teachers were asked to comment on why the recent changes in the time series standard have occurred. Reasons suggested were availability of new software (3), but more commonly others (10) suggested it was part of a move away from calculations involving time series and towards more interpretation of time series. Teacher 5 commented that the new standard, "allows students to tell the story of the data in a more real and authentic way". Teacher 2 related that

> for many students it [AS 3.1] had become a procedural exercise with EXCEL whereby they could learn the procedure for producing a forecast but had no idea what they had actually calculated in terms of context and why they were doing it…..They were not actually engaging with the context or doing any statistical thinking.

Seven teachers did not understand why the changes had occurred.

The next series of questions (Table 15) sought to discover whether resources used for teaching AS 3.1 and AS 3.8 had changed or remained largely the same.

Table 15 Comparison of different resources used or provided in the teaching of AS 3.1 and AS 3.8.

		Frequency	
	Description	**AS 3.1**	**AS 3.8**
Textbooks	Barton	7	0
	Nulake	4	1
	School produced Handbook	11	14
	None	1	3
Software	EXCEL	18	2
	iNZight	0	18
Hardware	Calculators/Graphics Calculators	8/13	7/13
	Computer Labs	15	13
	Tablets/Laptops	6	10
Context provided	None or less than 1 paragraph	4	1
	1 -2 paragraphs	7	6
	1 page	4	8
	2 pages	3	3
Links/References provided	Internet sites	3	12
	Articles	1	4
	Students do own research	4	12
	None	12	0

A number of substantial changes are revealed in Table 15. Firstly most schools have abandoned text-books when teaching AS 3.8, largely because those available have not proved useful. Teacher 3, when talking about obstacles, mentioned the "lack of published resources – Nulake not good, nor others". All the schools surveyed had moved from using EXCEL to *iNZight* although two still used EXCEL as well as *iNZight*. This represents a remarkable shift in the use of software, given that technological change, particularly in school curricula is notoriously slow (Oates, 2009) yet this shift has occurred in just a few years. Many schools (13) still rely on computer labs in order to access *iNZight* and the internet, but there was a slight increase in the use of tablets/laptops. Several teachers mentioned that their schools were intending to introduce a BYOD policy so it is anticipated that this shift will continue in the next few years. The amount of contextual content provided to students about time series data sets has increased from just a paragraph to more commonly a whole page. Eleven teachers noted that under AS 3.1 no or little contextual information was provided. Only four teachers provided contextual information in the form of websites and articles for AS 3.1 but sixteen teachers provided such information under AS 3.8. Under AS 3.1, only four teachers reported an expectation that students would do their own research, whereas under AS 3.8, twelve teachers expected students to investigate context by undertaking their own research, suggesting an important shift in this aspect.

Question 6 in part IV of the questionnaire asked teachers to compare how long they spent teaching certain aspects of the time series curriculum when teaching AS 3.1 as opposed to AS 3.8 (Table 16). Many teachers found this section of the questionnaire difficult to complete and given the range of totals it became clear that they were not all using the same units of measurement, despite the survey asking for measurement in terms of lessons or fractions of lessons. One teacher, for example, recalls 64 lessons for AS 3.1 and 78 for AS 3.8, whereas another recalls 15 for AS 3.1 and 17 for AS 3.8. So instead of comparing lesson or fractions of lessons numbers I decided to compare which topics teachers recorded as spending more or less time teaching when they reflected on their teaching under the two different standards.

Table 16 Comparison of teaching time spent on time series topics under AS 3.1 and AS 3.8.

Topic	More time under AS 3.8	More time under AS 3.1	No change
Features of time series	5	2	8
Exploratory activities	9	0	3
Additive models	2	6	5
Multiplicative models	4	4	5
Non-linear/piecewise trends	0	5	7
Indexed time series	0	4	7
Smoothing techniques in general	0	12	0
Trend description & quantification	5	1	7
Calculation of specific time series values	1	7	3
Calculation of seasonal effects	1	5	7
Calculation & interpretation of seasonally adjusted values	0	11	3
Learning to use visualisation software	5	5	4
Use visualisation software for data display production	7	2	4
Interpretation of data displays	6	0	5
Comparison with related time series	11	0	3
Production & interpretation of predictions	4	5	4
Evaluation of model	4	0	10
Research about context of time series	14	0	0
Communication of findings writing reports	10	0	4

Most of the results illustrated in Table 16 were as expected with more teaching time under AS 3.1 spent on smoothing techniques, non-linear and piecewise models and calculation of specific time series values, seasonal effects and seasonally adjusted values. Whereas under AS 3.8 more teaching time was devoted to exploratory activities, interpretation of predictions, comparison with similar time series, research about context and writing reports. The only perhaps surprising result is the category – learning to use visualisation software. Results show five teachers taking longer to teach it under AS 3.8, five taking longer under AS 3.1 and four reporting no change, or in other words quite a mixed experience of moving to *iNZight* software. This does not reflect the picture presented by the interviews which was that *iNZight* has a much shorter learning curve for students and teachers than EXCEL. For example, Teacher 1 noted, "[*iNZight*] is easy, easy, easy" and added "[under AS 3.1] I seemed to spend a lot of time teaching the mechanics of EXCEL – how to input formulae/create formulae". Moreover, the huge uptake of all 18 teachers in the use of *iNZight* is unlikely to have occurred if it was a difficult package to use. A possible explanation of this mixed reaction may be a consequence of familiarity with EXCEL compared to the unfamiliarity of *iNZight*. In conclusion to this section of the survey, sixteen teachers reported that their teaching of time series was either somewhat different or very different under AS 3.8 when compared with AS 3.1. Eight teachers had found implementation of the new standard manageable but five had found it somewhat difficult and four had found it quite easy.

The final section of the survey asked teachers for details of how the assessment process had changed between the two standards. One often mentioned challenge for teachers has been sourcing good time series data sets which contain enough interesting features to give students plenty to write about. Teachers have obviously been extremely resourceful in seeking out such time series but at the cost of hours of research when so many teachers are time-poor. Several teachers reported that this task was left to their Teacher-in-Charge representing a huge workload for such teachers since twelve respondents described this task as very difficult or somewhat difficult. Sources for data sets included – Statistics NZ (7), manufactured data sets (5), *iNZight* data files (4), NZAMT (4), Census@School (3), Google (2), TKI (1), colleagues and other schools (1). Similar sources were used to find contextual information to support the data sets.

As to the assessment itself, ten teachers reported that students in general found AS 3.8 more enjoyable than AS 3.1 although it was not without challenges. Four teachers reported poor student response to the new AS 3.8 standard, which they said students found to be too time-consuming and the writing of reports too difficult. There was a wide variation in how the assessment was managed by schools which some reported as a specific concern.

Teacher 2 comments:

> There is still an issue with assessments not being a level playing field….. we still do exam conditions for assessment to ensure authenticity of student work. Other schools allow take-home projects [but] we know tutors would give too much help if we allowed this.

This is confirmed by the survey results with only two teachers reporting that AS 3.8 is completed in class time only, whereas fifteen teachers reported that AS 3.8 at their school was completed as a mixture of class and students' own time. This also represents a big change from AS 3.1, when fourteen teachers reported that the assessment was completed in class time only and only three allowed a mixture of class and own time. Only one teacher reported that the assessment was completely conducted by students in their own time although other teachers mentioned that their school was moving in this direction. Teachers from one school noted that the assessment was conducted completely under exam conditions.

Teacher 11 mentioned that "marking [under AS 3.8] is overwhelming, a huge work-load and difficult to be consistent and fair". With this in mind it was interesting to see that the majority of teachers did not provide a page limit for assessment reports either under AS 3.1 or AS 3.8, although one teacher did mention that they were planning to introduce a page limit in order to try and contain the marking work-load.

Part V of the questionnaire sought information on the assessment process itself, including how it was managed and also asked teachers to reflect on differences in the quality and levels of reasoning and understanding demonstrated by students in the two standards.

Predominantly, Table 17 shows an increase in the perceived quality and level of understanding in student responses under the new standard when compared to the old, with thirteen teachers rating student work as much or slightly improved, sixteen teachers noting improvement in the integration of contextual information, twelve reporting a much higher or somewhat higher level of statistical reasoning and thirteen teachers reporting a much higher or somewhat higher level of understanding of time series. This supports anecdotal evidence that the quality of student exemplars under the new standard is higher.

One teacher's perception (Teacher 4) was that the level of statistical reasoning in AS 3.8 was somewhat lower than in AS 3.1. Another teacher's perception was that the level of understanding of time series was somewhat lower in AS 3.1 (Teacher 17). Their evidence in both cases was similar.

> Producing the smoothed data, calculating the seasonal effects etc. gave students a good understanding of features. When this is all presented for discussion there are a lot of misconceptions. (Teacher 4)

> Students may be better able to analyse the output from *iNZight* but they do not have the understanding of where the graph has come from. Using *iNZight* to produce the time series graph has reduced the possibility available to students to analyse the shape of the data and to consider alternatives (Teacher 17)

Anecdotal reports also confirm some teachers' scepticism about the effects on understanding of the *black-box* approach of *iNZight*.

Table 17 Perceived changes in student responses AS 3.8 compared to AS 3.1[2]

	Much improved	Slightly improved	No difference	Somewhat lower	Much lower
Quality of student work, 3.8 vs 3.1	9	4	2	1	0
Integration of contextual information, 3.8 vs 3.1	10	6	2	0	0
	Much higher	**Somewhat higher**	**No difference**	**Somewhat lower**	**Much lower**
Level of statistical reasoning, 3.8 vs 3.1	3	9	4	1	0
Level of understanding of time series, 3.8 vs 3.1	2	11	4	1	0

2. Some participants did not respond to these questions.

Finally teachers were invited to make any additional comments they wished concerning the teaching and learning of time series under the two different standards. Some of these comments clearly favoured the new standard, AS 3.8 with remarks like:

> [AS 3.8] gets students to think more about where the data has come from (Teacher 17)

> The focus has moved from creation of the time series graph [& underpinning values]….. to an understanding of the dataset itself (Teacher 17)

> It is harder to get away with it [rote learning under AS 3.8] as they have to write heaps more (Teacher 11)

> ….. emphasising context helps students understand influences on time series better (Teacher 1)

However, others felt

> calculation of trends, seasonal effects etc. helped student understanding better (Teacher 4)
>
> report writing and research was very challenging for some weaker students (Teacher 2)
>
> resourcing of AS 3.8 has been too time consuming (Teacher 3)

And Teacher 4 commented that it was

> hard for students to make meaningful statements about concepts that they have not explicitly been taught.

6.4 Teacher interviews

Information regarding the changes in the teaching and learning of time series was also investigated via teacher interview. Five teachers were interviewed, all of whom were lead or senior teachers of statistics at their schools. On some issues there was considerable agreement on others less so. In the main their responses supported those obtained from the survey results, any differences will be highlighted.

Teachers were initially asked about the merits of introducing time series into the secondary school curriculum. It is rare for time series to be explicitly taught at this level, it is usually the domain of undergraduate courses. However, the teachers interviewed were all very positive about its introduction into the NZ curriculum, mentioning the prevalence of time series in the media and the benefits of exposure to such data sets at school first. They were also all in agreement that of the new statistics standards introduced AS 3.8 was one of the more straightforward changes to manage. All teachers had attended professional development events in order to prepare for the new standard, citing events similar to those listed in Table 13.

Many of the new standards involved teachers using technology in their teaching in a different way. A range of technology was available to the teachers interviewed, from a very high tech school with each student having their own laptop, online learning systems, blogs, wikis and a library that established research pages by topic, to schools which still relied on booking computer labs to provide any access to technology for their students. All teachers discussed the value of having computer access available on demand when teaching Year 13 statistics. The availability of technology impacted teaching methods – some described having to ‘save up’ activities for the computer lab, whilst those with continual access tended to use technology more frequently but for shorter periods. Technology was used for research but also for access to *iNZight*. As reinforced by the survey data, all teachers interviewed had abandoned EXCEL, used in AS 3.1, for *iNZight* when teaching AS 3.8. Teacher 1 talked of her relief to no longer have to teach EXCEL skills, and described *iNZight* as “brilliant, easy, easy, easy!” However, others had found using *iNZight* a frustrating experience, noting downloading issues, sensitivity to data formats and lack of control over the scale of graphs as significant problems. MAC users, in particular, appeared to have more downloading problems than others. Teacher 1 also commented that "students have a shorter learning curve [with *iNZight*] than EXCEL". This comment was not verified by the survey data. Two schools got their students to download *iNZight* at home, with one of these schools requiring all graph production to be done at home.

Implementation of the new standard followed fairly similar paths at all five schools, with a lead teacher shouldering much of the burden at each school. It was the lead teacher who sourced interesting time series data sets, found contextual information and set assessments. In addition the lead teacher was responsible for the professional development of colleagues in the topic. Only one school pursued a different model with two teachers allocated to each standard; one writes the assessment including finding resources, the other acts as a moderator. Four of the five teachers assessed AS 3.1 under exam conditions during class or after school time. Only one school still retained this approach for AS 3.8; all the others had moved to a more project-style assessment with some class time allocated for working on the assessment, but the majority of it being completed in a student's own time. This change in the style of assessment has led to several teachers indicating that the time taken to complete this assessment is, under AS 3.8, considerably longer than it was under AS 3.1. However, one teacher thought the time factor was similar under both standards and another thought Excellence level students found the workload similar but lower ability students were finding the workload greater. They were all in agreement, though, that the marking workload represented a substantial increase for teachers.

Teachers also agreed that generally students find the time series standard one of the easier ones, but they couldn't agree on which standard students found harder – responses included probability, bivariate and inference. The literary component of the new standards was described by Teacher 2 as "very challenging" particularly for low ability students. Teaching of this literary component was tackled in a variety of ways; one school introduced literary structures from year 9 onwards, using writing frames, mnemonics and modelling good report writing, so by year 13 students' literary skills were largely well-developed. Another school produced exemplars illustrating overall structure and content of reports. An initial concern of this approach was that exemplars would be blindly reproduced or rote-learned, but they have found this not to be the case in practice. Two schools found peer reviewing of reports very effective in improving literary levels. One teacher commented that even under AS 3.1 they had encouraged their students to write full reports, but the students soon worked out that reports were not necessary to pass; this school is appreciative that the new standard supports their previous stance with comprehensive reports now being considered essential.

Teachers were asked to reflect on changes in their teaching approach with many discussing the almost total integration of technology into the teaching and assessment of time series as the biggest change. All also mentioned the increased emphasis on the context of the time series with Teacher 4 noticing her students “get a more in-depth knowledge rather than a knowledge that is gone as soon as the test is over”. Teacher 5 spoke about how “it’s nice to be able to get the kids thinking rather than just crunch numbers and get annoyed because they drag a cell down too far”. Although the teachers appeared to prefer the teaching, but not the marking, of the new standard, they all agreed that their students would probably prefer the old standard, particularly Achieved level or lazy students and those for who English was a second language.

When invited to make any final comments, one teacher expressed frustration about the moderation process. Her internal assessment had been returned by the moderator with a good score but with absolutely no written feedback which she had been counting on in order to provide direction for the future.

6.5 Summary of results

Student exemplars

- There has been a shift to a higher overall level of reasoning demonstrated in responses to AS 3.8 when compared with responses to AS 3.1. This shift is observed at all levels of achievement but is particularly noticeable at the Excellence level.
- There has been a substantial increase in the quantity of literary statements made in student exemplars for AS 3.8. This has resulted in an entirely different style of response from the calculation-dominated student responses of AS 3.1.
- Changes in the types of reasoning demonstrated at overall levels of reasoning have also been observed. At level 3, procedural reasoning, levels of reasoning shown in AS 3.8 exemplars are evenly spread across six sub-level categories,(3.1 to 3.6), but in AS 3.1 exemplars only four sub-level codes feature regularly (3.1, 3.3, 3.4 and 3.5).
- The shift to higher levels of reasoning in AS 3.8 is also demonstrated by achievement level – Achieved, Merit and Excellence. Only student exemplars from AS 3.8 demonstrate levels of reasoning at the interrogative level of reasoning. A few examples of interpretive reasoning were found in AS 3.1 exemplars but far greater proportions were found in AS 3.8 exemplars.

- At the three achievement levels, the levels of reasoning found in AS 3.8 exemplars demonstrated greater diversity in levels of reasoning than exemplars from AS 3.1.

Teacher Perceptions

- Teachers found professional development at Statistics Days the most useful in their preparation for the new standard. Opinion was mixed over whether they had felt well-prepared for the new standard or not.
- The software of choice for the new standard was *iNZight* with *all* teachers surveyed now using it. Two schools still used EXCEL as well. Mixed results were obtained regarding the amount of teaching time required to instruct students in the use of *iNZight*.
- The majority of teachers surveyed reported that the quality of student work under AS 3.8 was either much or slightly improved. Only one teacher thought the quality was lower.
- Introduction of the new standard had presented many challenges to teachers including lack of data-sets, lack of text-book resources, assessment management issues, increased marking workload. Lack of computing facilities no longer appears to be a major problem.
- Assessment of AS 3.8 was mainly conducted as a mixture of class and students' own time, whereas AS 3.1 was mainly conducted during class time only.
- Students' responses, in the form of a PPDAC-style report, were described as a steep learning curve for many, but particularly for weaker students.
- Teachers perceived an increase in teaching time spent on exploratory activities, interpretation of predictions, comparison with similar time series, research and report writing under AS 3.8. More time was spent on smoothing techniques, non-linear and piecewise models, and calculation of seasonal effects and seasonally adjusted values under AS 3.1.
- Many teachers reported that their students enjoyed the greater connection with the context of the data under AS 3.8. However, some teachers felt that the *black-box* approach of *iNZight* which automated the calculations required under AS 3.1 had been detrimental to student understanding.

CHAPTER SEVEN

Discussion and conclusions

7.1 Introduction

A brief history of the statistics curriculum in New Zealand, described in Chapter Two, provided background information surrounding the curriculum change which is the focus of this research. The curriculum change investigated is only one of many changes experienced by teachers in New Zealand so comments made by them may have been influenced by their overall perceptions of all changes as well as this specific change in the assessment of time series analysis. A particular factor in this curriculum change was the availability of data visualisation software. Integration of technology into teaching and assessment has been the subject of much research and it has been revealed to be a complex issue. The integration of *iNZight* into the teaching and assessment of time series was scrutinised in this research and factors important to its success discussed. A literature review of frameworks for the development of mathematical reasoning and dimensions of statistical reasoning were synthesised into a framework for the development of reasoning with time series. This synthesised framework, which was augmented based on student data, was used to determine the level of reasoning demonstrated in student exemplars completed before and after the time series curriculum change. Three types of data were collected in order to answer the main research question of whether learning outcomes have altered since the curriculum change. The data sources were from student exemplars of work, as well as surveys and interviews of teachers. This chapter discusses the three research questions, one main and two subsidiary questions. A number of pedagogical issues emerged from this research and these will also be discussed. Some recommendations for future research and for the professional development of statistics teachers are also made.

7.2 Main research question

In this section conclusions are drawn regarding the main research question. Conclusions are derived from the results of research conducted for this thesis and are related to findings from the literature. The main research question was:

> *How, if at all, have the learning outcomes for the new time series Achievement Standard, AS 3.8, changed when compared to the old Achievement Standard, AS 3.1?*

Student exemplars produced for the new standard, AS 3.8, demonstrated a higher quality and level of reasoning. This was shown through an analysis of over six hundred cases of reasoning levels found in student exemplars from both standards. Mean levels of reasoning calculated for exemplars at each achievement level confirmed a notable increase in the overall level of reasoning demonstrated by students had taken place. Such results were also confirmed by survey results which showed that thirteen teachers perceived a much improved or slightly improved quality of student work, with only one teacher's perception being that work was of a somewhat lower quality. In terms of the level of student reasoning, twelve teachers perceived much or slight improvement and again only one felt that the level of student reasoning was somewhat lower. A possible explanation for the perception of lower quality and level of reasoning is connected to the role that technology now plays. Three teachers felt that in AS 3.1 calculating the seasonal effects and smoothed series using EXCEL gave students a good understanding of the features of the time series. They added that in AS 3.8, where these calculations are no longer performed, students' misconceptions were more common. Research shows that when teachers demonstrate a preference for by-hand techniques and a lack of trust in technology, their attitudes are often passed on to their students (Stacey & Kendal, 2001). I maintain that this is a pedagogical issue and not a direct result of the introduction of the new standard or new technology. This and other pedagogical and technological issues will be discussed in Section 7.5. For many of the teachers, however, the shift away from EXCEL and consequently the need to teach EXCEL skills has liberated time for students to integrate their statistical and contextual knowledge at a deeper level. The dimensions of statistical reasoning developed by Wild and Pfannkuch (1999) include this element in dimension 2 and RME also stresses the importance of context and the use of real life situations (Heuvel-Panhuizen, 1996). One teacher (Teacher 17), commented

> The new standard certainly gets students to think more about where the data has come from and what the graphs means in the context of that dataset. However, it does require students to…..become experts in a field they do not know about. The focus has moved from being able to produce the time series graph and generally interpret it to an understanding of the dataset itself.

The integration of context is a crucial element of statistical reasoning and is a specific requirement of the new standard. Teacher 17's perceived difficulty of students having to become experts in a variety of fields mirrors the experience of practising statisticians and as such should be regarded as a positive learning outcome of the new Achievement Standard. One of the main catalysts for the recent changes in the statistics curriculum was to bridge the ever increasing gap between the thinking and practices of statisticians and the statistics that secondary school students were exposed to. An ability to *read behind the data* is the highest level of graphical reasoning identified by Shaughnessy (2007), which again emphasises the unique role of context in the discipline of statistics. The increase in the number of cases of interpretive reasoning was one of the biggest changes in learning outcomes between AS 3.8 and AS 3.1; further evidence that learning outcomes in AS 3.8 demonstrate higher levels of reasoning skills than AS 3.1.

Another change was the increase in cases of horizontal reasoning in AS 3.8 compared to AS 3.1. Horizontal reasoning is considered a lower level of reasoning than interpretive but has been the subject of much debate amongst teachers and has been discussed in some detail by the NZSA education committee. This increase in a lower level of reasoning appears to contradict other evidence that has shown a shift towards higher levels of reasoning. De Lange's pyramid (Figure 1) helps to explain this apparent paradox. The pyramid shows that lower level reproductive reasoning assessment tasks can range from simple to complex; horizontal reasoning fits into this category. As the pyramid visually demonstrates ability in performance of procedural type tasks, from simple to complex, is an important foundation for access to higher levels of reasoning. A time series whose overall trend is clearly linear represents a relatively simple task to describe. Under AS 3.1, such a scenario was common, a line of best fit was fitted to the smoothed linear series and the gradient used to describe the overall trend, a simple low level task. Without the scaffolding of the gradient, description of the trend has now become a more complex low level task. Under AS 3.8, and with other new statistics standards, teachers have been encouraged to use real data sets. Confronting students with these real data sets from a variety of contexts will never produce neat and tidy solutions; real data is messy but this is the reality for practising statisticians. This move towards using real data sets is another important change in the learning outcomes of the new standard.

A visual examination of student exemplars produced for the old and new standards provides the most obvious change in learning outcomes. The style of the new exemplars is that of a report which largely follows the PPDAC cycle, one of the dimensions of statistical thinking described by Wild and Pfannkuch (1999). This report style results in a greater number of written statements in exemplars as reports contain an introduction, analysis of results, discussion and conclusion. Exemplars produced for the old standard more commonly presented a series of calculations, often from an EXCEL spreadsheet, with one or two sentences on features of the time series such as long term trend. Another difference is the number and type of data visualisations included in the exemplars from the new standard; graphs are included of not only the raw data but also graphs of the components of the time series such as smoothed data, seasonal effects, residuals and predictions. Using different representations of a data set in order to improve understanding of that data set is termed *transnumeration* by Wild and Pfannkuch (1999, p 227).

They define transnumeration as

> …numeracy transformations made to facilitate understanding. Transnumeration occurs when we find ways of obtaining data…that capture meaningful elements of the real system. It pervades all statistical data analysis, occurring every time we change our way of looking at the data in the hope that this will convey new meaning to us.

Examination of new representations of the time series data set provided by *iNZight* is another important learning outcome change between the old and new standards.

7.3 Second research question

The second research question aimed to investigate and compare the types and levels of reasoning demonstrated in the two standards. The main research question confirmed that an increase in the level of reasoning had occurred, but this second research question investigated changes in sub-levels of reasoning as well as types and levels of reasoning observed at a variety of achievement levels.

Exemplars from the old Achievement Standard predominately (80%) displayed levels of reasoning at the procedural or extended procedural level, with around 10 per cent of cases from the horizontal and 10 per cent from interpretive levels of reasoning. This contrasts with just under 50 per cent of cases from the procedural levels of reasoning in exemplars from the new Achievement Standard. The percentage of cases at the interpretive level of reasoning *doubled* in the new Achievement Standard. Examples of the highest level of reasoning, interrogative reasoning was *only found* in exemplars from AS 3.8. The demand for more interpretive reasoning has meant that teachers have had to adjust how this standard is assessed, since interpretive reasoning takes time particularly when the context is unfamiliar. De Lange (1992) suggests that assessments which allow freedom of response are required to access more complex or higher levels of reasoning. Such open ended assessments, he suggests, are often managed as take-home tasks to allow students time to formulate responses but shorter in-class versions he maintains are also feasible. Most teachers reported that this assessment is now conducted partly in supervised class time and partly in students' own time; this additional time has provided the opportunity for students to access interpretive reasoning levels through research and further investigative activities, which was just not possible under the old standard given the time constraints of a supervised in-class assessment. Students accessing interrogative levels of reasoning also require additional time as evidenced by dimension 3, interrogative cycle, (Wild & Pfannkuch, 1999), which shows that this cycle may be repeated, possibly many times. A similar notion concerning the cyclical nature of the development of reasoning also appears in the Pirie-Kieren framework

(Pirie & Kieren, 1994) for the development of mathematical reasoning. In this framework the term *folding back* is used to represent re-examination of knowledge in order to progress to a higher level of reasoning. In the context of time series, a student may have pre-conceived ideas about factors determining the variation in a time series, but when they test these preconceptions in the light of further research and analysis they may require revision in order to deliver a valid interpretation of the data.

Changes in sub-levels of reasoning have also occurred, particularly at level 2, horizontal reasoning and level 3, procedural reasoning. A greater diversity of sub-codes in AS 3.8 was observed when compared with exemplars from AS 3.1. For example in level 3, AS 3.8 exemplars show roughly equal proportions (14%) in six different sub-codes, whereas in AS 3.1 exemplars, two sub-codes, 3.1 and 3.5 account for 56 per cent of all sub-codes at this level. A greater diversity in sub-codes was not observed at the higher levels of reasoning as few cases at these levels were found under AS 3.1.

Teachers' perception of changes in the level of reasoning was not as conclusive as the evidence provided by the student exemplars. Some teachers expressed concern that the lack of procedural computations in AS 3.8 was a retrograde step and appeared to rate procedural calculations as a higher order reasoning skill. This is not confirmed by any of the frameworks or dimensions used to produce the synthesised framework of Chapter Four. Procedural skills in terms of De Lange's pyramid (Figure 1) for example would be described as a lower level or reproduction skills. Typical procedural skills in the context of AS 3.1 included the reproduction of formulae in EXCEL in order to calculate smoothed series as well as seasonal effects. Within the SOLO taxonomy procedural skills such as calculation of smoothed series would be described as uni-structural, again one of the lower levels of reasoning described in this framework. Uni-structural tasks are defined as ones which require the student to focus on just one aspect of the task without having to relate this aspect to any other part of the task. Under AS 3.1, a student could focus on one set of calculations at a time without any reference to what those calculations meant in the context of the time series. The graphical output from *iNZight*, which many students use in AS 3.8, forces students to view several different representations of the data – raw, smoothed, residuals, seasonal effects – and to integrate the information from all representations in order to produce a meaningful description of the time series. I believe that teachers who hold on to the notion that procedural calculations represent higher order reasoning skills are those who have failed to grasp the critical difference between mathematical and statistical reasoning.

> …statistical reasoning is distinct from mathematical reasoning, as the former is inextricably linked to context. Reasoning in mathematics leads to discovery of mathematical patterns underlying the context, whereas statistical reasoning is necessarily dependent on data and context and requires integration of concrete and abstract ideas. (delMas, 2005, cited in Franklin, Bargagliotti, Case, Kader, Scheaffer & Spangler, 2015, p. 2)

This lack of understanding about the nature of statistical reasoning is hardly surprising given that many teachers of mathematics and statistics in New Zealand may have had little experience of data-driven statistical reasoning. Only a third of the teachers surveyed for this research had studied time series as part of their formal qualifications. If teachers have taken tertiary level statistics courses they are likely to have been theoretical and mathematically based. These results therefore highlight a need for greater professional development of teachers in order to cultivate their statistical understanding.

7.4 Third research question

In order to get a more holistic view of the changes that have occurred with the introduction of the new time series standard, data was also collected from teachers about their perception of the change. This data was collected via a survey and also from interviews of teachers. This data specifically assisted in answering the third research question which was:

> *What impact was there on teaching practice with the introduction of the new standard?*

Many teachers commented on the increased literary component required by AS 3.8. This was the most obvious difference when exemplars from the two standards were examined. One teacher had found teaching statistical report writing particularly challenging. To cope with this change teachers have employed a number of different strategies including:

- Establishing templates for reports, including PPDAC
- Peer review of reports
- Introduction of a simple report style at years 9 and 10, which is then refined to the level required for NCEA Level 3 by year 13.

Some teachers had resisted formulating templates as they suggested that this encouraged rote learning of key phrases, something that they said had been prevalent in AS 3.1. However, those that had used templates for AS 3.8 discovered that rote-learning had not been a problem in practice. Teachers who did not highlight any challenges with report writing had adopted a phased approach, introducing simple report styles in years 9 and 10, which could then be expanded and refined to reach the standard required by AS 3.8.

Use of appropriate technology is promoted throughout the new statistics standards and the new time series standard is no exception. All teachers reported using technology for both standards, but the shift from EXCEL to *iNZight* is quite remarkable in such a short time. All eighteen teachers reported that they used *iNZight* for AS 3.8, with two still using EXCEL as well. Oates (2009) identified a number of critical factors essential to the successful integration of technology into teaching and learning. These factors included organisational factors, access to technology, staff engagement, pedagogical content knowledge as well as integration of technology into all aspects of teaching, learning and assessment. Such critical factors appear to have been addressed with the introduction of *iNZight* enabling teachers in this study to shift their practice and change to *iNZight* technology. Teacher 6 commented "Use of *iNZight* has been accepted quickly and simply by students – more time can be spent on analysis". There was some contradictory evidence of the ease of uptake of *iNZight* with five teachers' perception that they had spent more time teaching students how to use *iNZight* than they had previously spent teaching students how to use EXCEL. This may in part be explained by the fact that when the data was collected it would have been the first time teachers as well as students had been exposed to *iNZight*. I predict that in future there would be very few teachers spending more teaching time on *iNZight* skills than they did on EXCEL skills as familiarity with the new software was achieved. Several teachers spoke of relief at no longer having to teach EXCEL skills any more. Teacher 1 recalls that under AS 3.1 "[I] seemed to spend a lot of time teaching the mechanics of EXCEL – how to input formulae/create formulae and then how to smooth etc.".

Although all teachers in this research now used technology to teach and assess this standard, teachers' access to technology varied considerably. Some teachers taught at schools where all students had lap-tops, whereas others relied on the availability of computer rooms or mobile banks of computers. Other schools had or were intending to institute a BYOD policy which would ensure availability of technology at all times as long as students remembered to bring their device to school. Those who needed to book computer rooms described having to "save up" activities for the computer room, which if given the choice they may have taught at a different time. No teacher cited lack of technology as an obstacle to the introduction of the new standard although the variety in levels of access did produce different teaching strategies. This suggests that the learning culture in the schools involved in this survey have largely embraced the integration of technology to some extent; this change in culture Raiti (2007) suggests is the most critical factor if technology is to be successfully integrated.

Types of teaching activities and time spent on them also changed between the two standards. In AS 3.1 teachers perceived that more of their time had been spent on teaching smoothing techniques, non-linear and piecewise models, calculations of seasonal effects and seasonally adjusted values. Under AS 3.8 teaching of these calculation skills largely disappeared with more teaching time now spent on exploratory activities, interpretation of predictions, comparison with similar time series, research and report-writing.

Teachers also reported that textbooks have largely been abandoned for this new standard due to a lack of what they perceived to be useful publications. Most schools had developed their own resources for this standard in the form of a booklet or handout which might be produced in hard or soft copy. For teaching and assessment purposes teachers had sought their own real life data sets for AS 3.8 rather than rely on ones supplied by whichever publication they had used for AS 3.1. Several teachers mentioned that sourcing data sets had been a time-consuming task and often was the responsibility of only one teacher at the school, creating a substantial additional work load for these teachers.

Assessment tasks were also reported as changing with most teachers providing more contextual information about the data set than they had done for assessment tasks for AS 3.1. The new research component had also led to a wider variety in the management of the assessment task. Under AS 3.1 most teachers reported that they had conducted this solely in class under exam conditions. Under AS 3.8, the situation had changed with only one school retaining this style, others moving to a combination of in-class and own time assessment or to a completely own-time project style assessment. This lack of consistency has fuelled requests for clear direction from NZQA about internal examination conditions. Teacher 2 notes "we still do an exam conditions assessment to ensure authenticity of student work. Other schools allow take-home projects [which] we know tutors would give too much help with if we allowed this".

7.5 Other issues

Pedagogical issues

Three teachers mentioned that they felt calculation of smoothed data and seasonal effects improved student understanding. Since this aspect had been removed with the shift to *iNZight*, these teachers suggested that the level of understanding of their students had deteriorated. This is an important observation but one that can be resolved easily by a change in pedagogical practice. Strong research evidence (e.g. Arnold, Pfannkuch, Wild, Regan & Budgett, 2011) suggests that *by-hand* activities conducted before the introduction of new technology is beneficial to students. In the context of AS 3.8 teachers could teach a variety of simple smoothing techniques by hand, such as moving average, weighted moving average, and exponential smoothing prior to the investigation of smoothed time series data using *iNZight*. A short tutorial was developed as part of professional development workshops held prior to the introduction of AS 3.8 which covers these techniques by hand. This tutorial can be conducted quickly and easily using EXCEL or the table function of a graphics calculator. It appears that some teachers omitted this important component of the time series learning trajectory.

Assessment issues

Many teachers raised concerns about the assessment itself, most of these concerns were expressed about internal assessments in general rather than the time series assessment in particular. As mentioned above many schools have moved to an assessment that may be completed wholly or partly in the students' own time. The view was expressed (Teacher 2) that this type of assessment did not now represent a level playing field given the diversity of internal assessment conditions that can now be experienced.

The workload involved in marking internal assessments with a research component was described by several teachers as overwhelming. One teacher interviewed felt that the workload had *tripled*. In order to combat this some schools had introduced a page limit for reports but others were faced with marking very lengthy documents. The workload in providing resources for this assessment was also significant particularly given that text-books had largely been abandoned for this topic. Often one lead teacher had spent many hours sourcing time series data sets with appropriate contexts and with enough features for students to have plenty to write about. Teachers were also sourcing more contextual information about the time series which yet again added to the overall work load, particularly that of the lead teacher. Some schools have begun to tackle this work-load by considering alternative forms of assessment. For example, one school is assessing AS 3.8 using a presentation rather than a written report.

Lack of moderation feedback was also mentioned as a problem. One teacher valued feedback from moderators for reassurance that their internal assessment was appropriate; feedback that was just a number without any comment was not considered to be helpful even if your number represented a good score. One teacher had found the grade boundaries harder to distinguish in AS 3.8 than they had been in AS 3.1. Greater clarity in AS 3.8 boundaries was suggested prior to the change in standard (Passmore, 2012b) but was not enacted. The proposal was to exploit the new interpretation opportunities offered by *iNZight* to establish these boundaries. In brief the proposal was for:

Achieved   Analysis of overall trend and features of one time series in context.

Merit   Comparison of overall trends and features of two related time series in context.

Excellence   Full analysis of two or more time series in context with reference to other research.

Technological issues

No issues with the implementation of *iNZight* were mentioned by teachers who participated in this research but other anecdotal evidence suggests that installation on MACs, at least initially, was more problematic. One teacher felt that the graphics on *iNZight* were poor, citing scaling and an inability to read data values accurately. Since this research was conducted a number of improvements have been made and are still being made to *iNZight* which I suggest have resolved the majority of issues raised by early users of *iNZight*.

Access to computing facilities was available in all schools who participated in this research, but with varying degrees of availability. At some schools students brought lap-tops to every class, other schools operated a BYOD policy while some schools still relied on booking computer labs or mobile sets of laptops. This latter group acknowledged that permanent availability of access to technology for their students would make the teaching of this and the other Level 3 statistics standards much easier. As all the teachers who participated in this research taught at decile 7 or higher schools, the fact that no issues with technology access were raised is perhaps not surprising. It would be interesting to discover whether technology access is also no longer an issue for teachers at low decile schools.

Research shows that when teachers demonstrate a preference for by-hand techniques and a lack of trust in technology, their attitudes are often passed on to their students (Stacey & Kendal, 2001). This lack of trust in technology was evident in the responses of some participants in this research. Goos et al. (2000) shows that there are many types of interactions between students, teachers and technology. The teachers expressing mistrust of technology are uncomfortable with the blind acceptance of the output of technology by their students. Such teachers have yet to reach the level of interaction described in the research as *technology as a partner* or *technology as an extension of self*. Only when these levels are reached do the true learning benefits of technology emerge.

Rationale for curriculum changes

The curriculum change which is the focus of this research is only one of a raft of changes that have been introduced since 2010. Seven of the teachers surveyed were unsure about why these changes had occurred. It is disappointing that the catalyst behind a curriculum change as significant as the one seen in the New Zealand statistics curriculum has not been adequately conveyed to the wider teaching community. The reforms are underpinned by substantial research including two particularly influential reports; the GAISE report (2005) and George Cobb's paper, The introductory statistics course: A ptolemaic curriculum (2007). The teachers in this research were selected from lists of schools supporting professional development events. It is a concern that the message regarding the curriculum change may have had an even weaker exposure in schools that did not support such professional development events. Without an understanding of the rationale behind the curriculum changes teachers could be more reluctant to embrace those changes, viewing them as simply additional work load rather than an opportunity to improve student learning outcomes. The goal of improving student reasoning and narrowing the gap between professional statistician practice and the statistics taught in NZ secondary schools has obvious benefits for todays' students that surely all professional teachers would wish to support if they were aware of it.

Statistical content knowledge of teachers

Under the old standard, AS 3.1, concern was expressed by Mike Camden (personal communication, 7 April 2015), NZSA Statistics Education Committee, that some students were able to pass this standard by implementing an incorrect statistical technique, in particular, fitting a linear trend to non-linear data.

> In the past, I've been alarmed to see the practice of fitting a straight line to the 'smooth'. At least that's better than fitting it to the raw data! … I'm pleased [this practice] has quietly gone.

And when referring to description of the time series he goes on to say

> We don't want them [students] describing each point. We want [the students] describing features (possibly explainable) that stand out from the noise (unexplainable).

The practice of fitting a linear trend to obviously non-linear data demonstrates a lack of statistical content knowledge of teachers. This problem is not unique to New Zealand, indeed the American Statistical Association published a document in April 2015, known as the SET document (Franklin et al., 2015) containing a whole raft of professional development recommendations to improve the content knowledge of teachers of statistics in America.

Three recommendations are made specifically for teachers of secondary school statistics, including:

- An introductory course emphasising a data-driven approach to statistical thinking.
- A second course focussing on re-randomisation, experimental design, inference and independence tests.
- A statistical modelling course.

A particular issue for teachers with the new standard, AS 3.8, demonstrated by the additional teaching time being spent on this area, is the description of the overall trend of a time series.

Under the old standard, scaffolding for this description was provided by using the gradient of a line of best fit fitted to the smoothed data. Students were taught to use the gradient value in a one sentence description of the trend, for example one student wrote, "*The gradient of the trend line shows that the sales of footwear are increasing by $0.59 million per quarter*". Students using *iNZight* to produce the smoothed data no longer have the scaffolding of the gradient to use as the basis for their description of overall trend but have to examine the smoothed data carefully to uncover this information. Since real data sets are being more commonly used whose smoothed series do not often exhibit neat linear trends, students now have to consider when any major changes in trend have occurred and learn to distinguish a short term deviation from a shift in overall trend. Such a task is straightforward for an experienced observer of time series but is proving to be a challenge for some students and teachers. Given these difficulties I suggest that a rule of thumb be developed to help students discount small deviations to the overall trend but comment only major changes in overall trend. Interpretation of CUSUM charts (Taylor, 2014) which are CUmulative SUMs of differences between the values and the average or investigation of consecutive values of positive/negative residuals could be utilised for the rule of thumb.

7.6 Implications for teaching, assessment and research

The synthesised framework (Table 7) used to analyse student exemplars produced for the old and new standards could also be used as a framework for teaching units in the topic of time series. It can help teachers to create a unit that starts with exposing students to activities which require a lower level of reasoning such as overall trend description. Having mastered this level of reasoning students can then be guided towards the higher levels of reasoning such as interpretive and interrogative. The process of creating a synthesised framework has already been used in other areas of statistical education research (e.g., Mooney et al., 2014) but this research helps to confirm the efficacy of such an approach and will hopefully encourage other researchers to attempt the creation of other useful frameworks.

Teacher participants in this research raised a number of largely administrative concerns regarding not only the assessment of time series but internal assessment and moderation in general. Teachers require some guidance about expectations in order that their students are fairly assessed in comparison to other students in New Zealand. Unless such guidance is forthcoming inequities in internal assessment will persist thereby diminishing the value of such assessments to future employers or future education providers.

The release of *iNZight* has irrevocably changed the way statistics is taught by the participants in this research, but it is disappointing that some of the excellent features of the time series module have not been exploited by the new standard. In particular, it is very easy to compare and contrast related time series within *iNZight* and this provides useful insight into the context of the data sets and can assist in the interpretation of features. Lack of clarity in grade boundaries in the new standard is mentioned as an issue by some teachers, and this feature of *iNZight* is an obvious and straightforward opportunity to amend the grade boundaries by adding 'comparison of related time series' to the criteria required for a Merit grade.

The topic of time series in many countries is restricted to undergraduate courses and often secondary school curricula do not focus on this topic. As a result international comparisons have not formed part of this research, but future research in this area could investigate how time series is taught globally; comparing and contrasting student learning outcomes. The results of this research also indicate that *all* curriculum changes in New Zealand could benefit from a similar scrutiny in order to examine any changes in student learning outcomes. Similar synthesised frameworks could be developed against which student exemplars are analysed in order to categorically answer the question as to whether the changes in the New Zealand statistics curriculum has produced higher levels of reasoning in student learning outcomes. An inhibitor of such research is the difficulty in accessing student work; ethical considerations often mean this rich data source is unattainable.

Finally this research has highlighted a need for greater professional development of statistics teachers if important curriculum change is to be embraced successfully by all teachers. The changes in the statistics curriculum are so significant that the topics of statistics and mathematics now demand separate and distinct skill sets. Without an appreciation of the differences in statistical and mathematical reasoning for example, students can be exposed to a whole raft of misconceptions, and emerge from school ill-equipped to deal with tertiary level statistics courses.

7.7 Limitations

The research methods employed and described in Chapter Five were not those envisaged for an ideal research design, thus any user of the results of this research should be cognisant of their limitations. The primary data source was student exemplars circulated by NZQA. Although these are exemplars of actual student work they were not a randomly selected representative sample of student work. Problems of access to actual student work prevented the selection of such a representative sample. Two further data sources were from questionnaires completed by teachers and interviews of teachers. The teachers involved did not constitute a representative sample but rather a *purposive sample*. The teachers were selected from schools which had participated in past professional development events; their views may differ markedly from those of teachers who do not attend professional development events. The quality of the data from teachers may also have been affected by questionnaire fatigue, indicated in some responses, and the interviews by interviewer bias as the researcher was known to many of the interviewees. The research also suffered from non-response bias with only 62% of the questionnaires returned completed and of the nine schools invited to participate, one declined. Therefore the findings from this study should be treated with caution.

7.8 Conclusion

This research sought to discover whether a curriculum change in the secondary school topic of time series had changed learning outcomes of students. In order to objectively analyse this change a framework was developed of the types and levels of reasoning in time series (Table 4). This framework was constructed as a synthesis of several other well-established frameworks. Ethical and time considerations meant that samples of student NCEA work before and after the curriculum change from a variety of schools could not be accessed. Instead exemplars of student work, distributed by NZQA for each Achievement Standard, and freely available on the internet were used. The analysis provided strong evidence that an overall shift towards higher levels of reasoning had taken place following the curriculum change. Within this overall shift an increase in examples of horizontal, a lower level of reasoning, was also observed. This shift appears to be the result of teachers using real data sets which has changed the low level simple task of overall trend description into a much more complex task. Another obvious change is the style of student responses, which has changed from a presentation of a series of calculations with some commentary to a more formal report style with an introduction followed by sections on results, analysis and conclusions. This curriculum change is only one of many introduced in the New Zealand statistics curriculum since 2010; further research is recommended to investigate whether similar changes in the learning outcomes of students have occurred in other Achievement Standards.

REFERENCES

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

Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands on to movies. *Journal of Statistics Education, 19*(2), 1-32.

Begg, A., & Pfannkuch, M. (2004). *The school statistics curriculum : statistics and probability education literature review.* University of Auckland: Auckland Uniservices Ltd.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman, & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistics Institute.

Berger, M. (1998). Graphics calculators: An interpretative framework. *For the Learning of Mathematics, 18*(2), 13-20.

Biehler, R., Ben-Zvi, D., Bakkar, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements (Ed.), *Third international handbook of mathematics education* (pp. 643-689). New York, NY: Springer.

Borg, W. R. (1963). *Educational research: A practical guide for teachers.* New York, NY: Longman.

Brookhart, S. M. (2010). *How to assess higher order thinking skills in your classroom.* Alexandria, VA: ASCD.

Camden, M. (2003). *Time series in AS92580*. Retrieved from New Zealand Statistics Association: http://www.stats.org.nz/Newsletter58/StatIn%20NZSchools.doc

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), 1-26.

Clegg, B. (2007). Qualitative data from a postal questionnaire: Questioning the presumption of the value of presence. *International Journal of Social Research Methodology, 10*(4), 307-317.

Cobb, G. (2007). The Introductory Statistics Course : A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1 - 15.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education.* Abingdon, UK: Routledge.

Collis, K., & Biggs, J. (1982). *Evaluating the quality of learning - the SOLO taxonomy.* New York, NY: Academic Press.

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage Publications.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393.

de Lange, J. (1992). Assessment: No change without problems. In M. Stephens, & J. Izard (Eds.), *Proceedings from the First National Conference on Assessment in the Mathematical Sciences* (pp. 1-28). Geelong, Australia: The Australian Council for Educational Research Ltd.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7,* 1-26.

Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association, 95*(452), 1293-1296.

Erickson, T. (2002). *Fifty fathoms: Statistics demonstrations for deeper understanding.* Oakland, CA: EEPS Media.

Finzer, W. (2006). *Fathom dynamic data software.* Emeryville, CA: Key Curriculum Press.

Forbes, S. (2014). The coming of age of statistics education in New Zealand, and its influence internationally. *Journal of Statistics Education, 22*(2).

Forbes, S., Chapman, J., Harraway, J., Stirling, D., & Wild, C. (2014). Use of data visualisation in the teaching of statistics: A New Zealand perspective. *Statistics Education Research Journal, 13*(2), 187 - 201.

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *The statistical education of teachers.* Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/SET/SET.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics (GAISE) report*. Alexandria, VA: American Statistical Association. Retrieved from American Statistical Association: http://www.amstat.org/education/gaise/gaiseprek-12_full.pdf

Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124-158.

Gal, I., & Garfield, J. (Eds.). (1997). *The assessment challenges in statistics education.* Amsterdam, The Netherlands: International Statistical Institute.

Gardner, P. (2003). Oral history in education: teachers' memory and teachers' history. *History of Education, 32*(2), 178-188.

Goos, M., Galbraith, P., Renshaw, P., & Geiger, V. (2000). Reshaping teacher and student roles in technology-enriched classrooms. *Mathematics Education Research Journal, 12*(3), 303-320.

Handal, B., & Herrington, A. (2003). Mathematics teachers' beliefs and curriculum reform. *Mathematics Education Research Journal, 15*(1), 59-69.

Haynes, M. (1996). *Influences on practice in the mathematics classroom: An investigation into the beliefs and practices of beginning teachers.* Unpublished master's thesis. Palmerston North, NZ: Massey University.

Heuvel-Panhuizen, M. V. (1996). *Assessment and realistic mathematics education.* Utrecht, The Netherlands: CD -Beta Press/Freudenthal Institute.

Hillel, J. (1993). Computer algebra systems as cognitive technologies: Implication for the practice of mathematics education. In K. Ruthven, & C. Keitel (Eds.), *Learning from Computers: Mathematics Education and Technology* (pp. 18-47). Berlin, Germany: Springer-Verlag.

Hitchcock, G., & Hughes, D. (1995). *Research and the teacher.* London: Routledge.

Huberman, A., & Miles, M. (1994). *Data management and analysis methods*. Thousand Oaks, CA: Sage Publications.

Irwin, B., & Jones, N. (2014, January 25). Pack your laptop, we're off to school. *New Zealand Herald*. Auckland. Retrieved from http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11191631

Johnson, B., & Christensen, L. (2004). *Educational research - Quantitative, qualitative and mixed approaches.* Thousand Oaks, CA: Sage Publications.

Kelly, A., & Lesh, R. (2000). Trends and shifts in research methods. In A. Kelly, & R. Lesh (Eds.), *Hand book of research design in mathematics and science education* (pp. 35-44). Mahwah, NJ: Lawrence Erlbaum Associates.

Konold, C., & Miller, C. (2005). *Tinkerplots: Dynamic data explosion.* Emeryville, CA: Key Curriculum Press.

Konold, C., Higgins, T., Russell, S., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305-325.

Krosnick, J., & Presser, S. (2010). Question and questionnaire design. In P. Marsden, & J. Wright (Eds.), *Handbook of survey research* (pp. 263-313). Bingley, UK: Emerald Group Publishing Ltd.

Kvale, S. (1996). *Interviews.* Thousand Oaks, CA: Sage Publications.

Learning with 1:1 digital devices. (2014, July 28). Retrieved from Te Kete Ipurangi: http://elearning.tki.org.nz/Technologies/Learning-with-1-1-digital-devices

Lynch, J. (2006). Assessing effects of technology usage on mathematics learning. *Mathematics Education Research Journal, 18*(3), 29-43.

Martin, L. (2008). Folding back and the dynamical growth of mathematical understanding: Elaborating the Pirie-Kieren theory. *The Journal of Mathematical Behavior, 27*(1), 64-85.

Martin, P. (1993). *The VCE mathematics experiment: An evaluation.* Geelong, Australia: Deakin University National Centre for Research and Development in Mathematics Education.

Martin, W. O. (2000). Lasting effects of the integrated use of graphing technologies in pre-calculus mathematics. In E. Dubinsky, A. Schoenfeld, & J. Kaput (Eds.), *Research in Collegiate Mathematics Education. IV: Issues in Mathematics Education* (pp. 154-187). Washington, DC: American Mathematical Society.

McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics, 45*(1-3), 103-129.

Ministry of Education. (1992). *Mathematics in the New Zealand curriculum.* Wellington, New Zealand: Learning Media.

Ministry of Education. (2007). *The New Zealand curriculum.* Wellington, New Zealand: Learning Media.

Mooney, E. S., Langrall, C. W., & Hertel, J. T. (2014). A practitional perspective on probabilistic thinking models and frameworks. In E. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking* (pp. 495-507). Dordrecht, The Netherlands: Springer.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123- 165.

Murray Thomas, R. (2003). *Blending qualitative and quantitative research methods in theses and dissertations.* London, UK: Corwin Press.

New Zealand Qualifications Authority. (2013). *Mathematics and statistics 3.8.* Retrieved from New Zealand Qualifications Authority: http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2013/as91580.pdf

Oates, G. (2009). *Integrated technology in the undergraduate mathematics curriculum: A case study of computer algebra systems.* Auckland, NZ: University of Auckland. Retrieved from https://researchspace.auckland.ac.nz/handle/2292/4533

Olson, H., & Keynes, A. (2001). Professional development for changing undergraduate mathematics instruction. In D. Holton (Ed.), *The teaching and learning of mathematics at the university level: An ICMI study* (pp. 113-126). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Palinkas, L., Horowitz, S., Green, C., Wisdom, J., Duan, N., & Hoagwood, K. (2014). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533-544.

Passmore, R. (2012a). *A teacher's guide to Holt-Winters analysis.* Retrieved from Census@School: http://new.censusatschool.org.nz/resource/teachers-guide-to-holt-winters-analysis/

Passmore, R. (2012b). *Food for thought, possible student exemplars.* Retrieved from Census @ School: http://new.censusatschool.org.nz/resource/food-for-thought/

Patterson, J., & Czaijkowski, T. (1979). Implementions: Neglected phase in curriculum change. *Educational Leadership, 37*(3), 204 - 206.

Pea, R. (1985). Beyond amplification: Using the computer to reorganise mental functioning. *Educational Psychologist, 20*(4), 167-182.

Pegg, J. (1992). Assessing students' understanding at the primary and secondary level in the mathematical sciences. In M. Stephens & J. Izard (Eds.), *Reshaping assessment practice: Assessment in the mathematical sciences under challenge* (pp. 368-385). Melbourne, Australia: Australian Council of Educational Research.

Pegg, J. (2005). Assessment in mathematics: A developmental approach. In J. Royer (Ed.), *Mathematical Cognition* (pp. 227 - 259). Greenwich, CT: Information Age Publishing.

Petty, N. (2013, February 11). *EXCEL, SPSS, Minitab or R?* Retrieved from https://learnandteachstatistics.wordpress.com/tag/inzight/

Pfannkuch, M. (2008). Training teachers to develop statistical thinking. In C. Batanero (Ed.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.* Monterey, Mexico: International Commission on Mathematical Instruction and International Association for Statistics Education. Retrieved from http://iase-web.org/documents/papers/rt2008/T4P2_Pfannkuch.pdf

Pfannkuch, M. (2011, June). Statistics Education News. *New Zealand Statistical Association Newsletter 73*. Retrieved from http://stats.org.nz/Newsletter73/education.htm

Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 323-333). New York, NY: Springer.

Pirie, S., & Kieren, T. (1994). Growth in mathematical understanding: How can we characterise it and how can we represent it? *Educational Studies in Mathematics, 26*(2-3), 165-190.

Punch, K. (2009). *Introduction to research methods in education.* London, UK: Sage.

Raiti, J. (2007). It's about culture: How to really integrate ICT. *Teacher, 183,* 10-13.

Reason, P., & Bradbury, H. (Eds.). (2008). *The handbook of action research.* London, UK: Sage.

Resnick, L. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Romberg, T., Zarinnia, E., & Collis, K. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics* (pp. 21 -38). Washington, DC: American Association for the Advancement of Science.

Scott, J. (1990). *A matter of record, documentary sources in social research.* Cambridge, UK: Polity Press.

Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957-1009). Charlotte, NC: Information Age Publishers.

Smith, D. (1998). Renewal in collegiate mathematics education: Learning from research. *Documenta Mathematica, Extra Volume ICM 1998, 3*, 778-786.

Sokal, R., & Rohlf, F. (2012). *Biometry* (4th ed.). New York, NY: W H Freeman.

Stacey, M., & Kendal, K. (2001). The impact of teacher privileging on learning differentiation. *International Journal of Computers for Mathematical Learning, 6*(2), 143-165.

Stacey, R., & Pierce, K. (2001). Observations of students' responses to learning in a CAS environment. *Mathematics Education Research Journal*, *13*(1), 28-46.

Stake, R. (2010). *Qualitative research: Studying how things work.* New York, NY: Guilford Press.

Stigler, S. (1999). *Statistics on the table: The history of statistical concepts and methods.* Cambridge, MA: Harvard University Press.

Tall, D. (2000). Cognitive development in advanced mathematics using technology. *Mathematics Education Research Journal, 12*(3), 210-231.

Taylor, W. (2014, December 31). *Change-point analysis: A powerful new tool for detecting changes.* Retrieved from www.variation.com: http://www.variation.com/cpa/tech/changepoint.html

Thomas, M., & Chinnappan, M. (2000). Function representation and technology-enhanced teaching. In A. Chapman, & J. Bana (Ed.), *Twenty-Third Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 172-179). Perth, Australia: Mathematics Education Research Group of Australasia.

Tukey, J. (1977). *Exploratory data analysis.* Boston, MA: Addison-Wesley.

Verhage, H., & De Lange, J. (1997). Mathematics education and assessment. *Pythagoras*, 14-20.

Wallen, N., & Fraenkel, J. (2001). *Educational research. A guide to the process.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1.

Watson, J. M. (1994). A diagrammatic representation for studying problem-solving behavior. *Journal of Mathematical Behavior* (3), 305-332.

Wellington, J. (2000). *Educational research - Contemporary issues and practical approaches.* London, UK: Continuum.

Whyte, W. (Ed.). (1991). *Participatory action research.* Newbury Park, CA: Sage.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

APPENDIX 1

Questionnaire – Year 13 statistics teacher

NAME………………………………………………………………………………………………

Year 13 Statistics Teacher Questionnaire

Instructions

Circle or underline the appropriate response(s) to the following questions.

Part I: Background Information on School

1) **Type of School.**
   a. State
   b. Integrated
   c. Private
   d. Other please specify

2) **What is the decile ranking of your school? Please enter number between 1 and 10.**

3) **Which response best describes the location of your school?**
   a. Rural
   b. Inner Urban
   c. Outer Urban

4) **What is your school type?**
   a. Co-educational
   b. Boys only
   c. Girls only

5) **Does your school have a high proportion (over 40%) of students for whom English is their second language? Yes/No**

6) **How would you describe your Mathematics Department?**
   a. Very traditional
   b. Fairly traditional
   c. A mixture of traditional and progressive
   d. Fairly progressive
   e. Very progressive

Part II: Background Information about Teacher

1) **Position held in Mathematics Department**

   a. HOD
   b. Teacher in charge of Year 13 Statistics
   c. Other TIC of Mathematics Year Level
   d. Year 13 Statistics teacher

2) **Teacher qualification obtained in?** **NZ/ Overseas**

3) **How many years have you taught Year 13 Statistics?**

   a. First year
   b. 1 – 2 years
   c. 3 – 5 years
   d. 6 – 10 years
   e. Over 10 years

4) **How many years have you taught time series analysis?**

   a. First year
   b. 1 – 2 years
   c. 3 – 5 years
   d. 6 – 10 years
   e. Over 10 years

5) **Did you study time series as an undergraduate or post graduate student?**

   a. Undergraduate
   b. Post-graduate
   c. Neither

6) **Which time series Achievement Standards have you taught?**

   a. AS 3.8 only
   b. AS 3.1 only
   c. Both AS 3.1 and AS 3.8

Part III: Professional Development

1) Have you developed your knowledge of time series by any of the following means in the last two years to help

cope with teaching the new Achievement Standard 3.8 ?

School run professional development	*Yes/No*
Professional development run by local Maths association	*Yes/No*
Professional development run by a private provider	*Yes/No*
NZQA facilitated workshop	*Yes/No*
From colleagues at your own or another school	*Yes/No*
From Census@School online resources	*Yes/No*
From other on-line resources, please specify	*Yes/No*
Text-books	*Yes/No*
Self-directed study	*Yes/No*
Attendance at a conference eg. NZAMT	*Yes/No*
Attendance at a Statistics Teacher Day	*Yes/No*
Other, please specify	

2) Rank the sources listed above **that you used** in terms of usefulness? Rank 1 being the most useful.

School run professional development	*Rank………*
Professional development run by local Maths association	*Rank………*
Professional development run by a private provider	*Rank………*
NZQA facilitated workshop	*Rank………*
From colleagues at your own or another school	*Rank………*
From Census@School online resources	*Rank………*
From other on-line resources, please specify	*Rank………*
Text-books	*Rank………*
Self-directed study	*Rank………*
Attendance at a conference eg. NZAMT	*Rank………*
Attendance at a Statistics Teacher Day	*Rank………*
Other, please specify	*Rank………*

3) **Did you feel competent to teach the new time series Achievement Standard?**

a. Strongly Agree
b. Agree
c. Not Sure
d. Disagree
e. Strongly Disagree

4) **Did you feel well prepared to mark the new time series Achievement Standard?**

a. Strongly Agree
b. Agree
c. Not Sure
d. Disagree
e. Strongly Disagree

5) **Did you feel confident about teaching students how to structure and write their reports?**

a. Strongly Agree
b. Agree
c. Not Sure
d. Disagree
e. Strongly Disagree

6) **Are there any obstacles that have prevented you teaching time series in the way that you wish? Please specify.**

7) **How would you describe the workload of implementing AS 3.8?**

a. Substantial
b. Somewhat substantial
c. Not Sure
d. Somewhat light
e. Very light

8) **Are you aware of <u>WHY</u> the recent changes in the time series Achievement Standard have occurred?**

Yes/ No If yes, please describe the rationale behind these changes.

Part IV: Comparison of the teaching of AS 3.1 and AS 3.8

Questions	AS 3.1	AS 3.8
1) Which textbook did you use? Please circle.	*Barton, Cambridge, Nulake AME, Handbook produced by your school, None,* *Other please specify*	*Barton, Cambridge, Nulake, AME, Handbook produced by your school, None,* *Other please specify*
2) Which software package(s) did you use? Please circle.	*EXCEL, Fathom, Genstat, iNZight, SAS, Tinkerplots,* *None, Other please specify*	*EXCEL, Fathom, Genstat, iNZight, SAS, Tinkerplots,* *None, Other please specify*
3) What technological devices were available for your students to use when you were teaching ? Please circle.	*Scientific calculators, Graphics calculators, Computer lab, laptops/tablets/ipads,* *None, Other please specify*	*Scientific calculators, Graphics calculators, Computer lab, laptops/tablets/ipads,* *None, Other please specify*
4) How much specific contextual information did you provide for your students? Please circle.	*A paragraph or two, one page, two pages, three or more pages, no contextual information provided,* *other please specify*	*A paragraph or two, one page, two pages, three or more pages, no contextual information provided,* *other please specify*

5) Did you provide links or references to contextual information for your students? Please circle appropriate response(s).	*Links to internet sites, articles from journals/newspapers, book references, students expected to do own independent research*	*Links to internet sites, articles from journals/newspapers, book references, students expected to do own independent research*
Questions	**AS 3.1**	**AS 3.8**
6) In your typical teaching unit on time series, **roughly** how many lessons did you spend on the following:	**Number of lessons or fractions of lessons**	**Number of lessons or fractions of lessons**
Features of time series		
Investigative and exploratory activities		
Additive models		
Multiplicative models		
Non-linear or piecewise trends		
Indexed time series		
Smoothing techniques (Moving Means/Medians, Centred Moving Means/Medians,		

weighted moving means/medians, exponential, Holt-Winters, ARIMA , other). Please indicate those taught for each standard as well.		
Trend description and quantification		
Calculation of specific time series values		
Calculation and interpretation of seasonal effects		
Calculation and interpretation of seasonally adjusted values		
Question 6 continued	**AS 3.1(Number of lessons)**	**AS 3.8(Number of lessons)**
Learning how to use visualisation software		
Using visualisation software for production of data displays		
Interpretation of visual representations of time series		
Comparison with related time series		
Production and interpretation of predictions		
Evaluation of model		
Research about the context of the time series		
Communication of findings; writing reports		

7) **How similar is the way you teach AS 3.8 compared to how you used to teach AS 3.1?**

a. Very Similar
b. Somewhat Similar
c. Same
d. Somewhat different
e. Very different

8) **How have you found implementing AS 3.8 in the classroom?**

a. Very difficult
b. Somewhat difficult
c. OK
d. Quite easy
e. Very easy

Part V: The assessment process – AS 3.1 vs AS 3.8

1) **Where did you source datasets for assessment purposes?**

   a. NZQA exemplars
   b. *iNZight* data files
   c. Census@School
   d. Statistics NZ
   e. Datamarket
   f. Google search
   g. NZAMT
   h. TKI
   i. Own or other resources, please specify

2) **Describe the process of sourcing these datasets for internal assessment?**

   a. Very difficult
   b. Somewhat difficult
   c. OK
   d. Somewhat easy
   e. Very easy

3) **Where did you source contextual information for assessment purposes?**

   a. NZQA exemplars
   b. *iNZight* data files
   c. Census@School
   d. Statistics NZ
   e. Datamarket
   f. Google search
   g. NZAMT
   h. TKI
   i. Own or other resources, please specify

4) **How have your students responded to the new AS 3.8 assessment compared to AS 3.1?**

   a. Disliked it
   b. Not very keen, found some aspects very hard
   c. No difference to AS 3.1

d. Enjoyable but some challenges
e. Really liked it , coped well

5) **When did your students complete the assessment?**

AS 3.1

a. In class time only
b. A mixture of class time and own time
c. Own time only

AS 3.8

a. In class time only
b. A mixture of class time and own time
c. Own time only

6) **Did your students have a page limit for their assessment reports?**

a. AS 3.1 – No Limit, Limit of …………………… pages
b. AS 3.8 – No Limit, Limit of ……………………. pages

7) **Did your students have a choice about which time series to use for their assessment?**

a. AS 3.1 – no choice/choice from a few provided time series/ sourced own series
b. AS 3.8 – no choice/choice from a few provided time series/sourced own series

8) **In your opinion would you say the quality of student work produced for AS 3.8 when compared with AS 3.1 is:**

a. Much improved
b. Slightly improved
c. No difference
d. Somewhat lower
e. Much lower

9) **In your opinion would you say the level of student statistical reasoning illustrated in work produced for AS 3.8 when compared with AS 3.1 is:**

a. Much higher
b. Somewhat higher
c. No difference
d. Somewhat lower
e. Much lower

10) **In your opinion would you say that the ability of students to integrate contextual information into work produced for AS 3.8 when compared with work produced for AS 3.1 is:**

a. Much improved
b. Slight improvement
c. No difference
d. Somewhat worse
e. Much worse

11) **In your opinion would you say that the level of understanding of time series of students who studied AS 3.8 when compared with those who studied AS 3.1 is:**

a. Much higher
b. Somewhat higher
c. No difference
d. Somewhat lower
e. Much lower

Part VI: Any Other Comments

Please use this section to write any comments you may have about the teaching and learning of time series in Year 13 based on your own and your students' experience of the two Achievement Standards 3.1 and 3.8.

Thank you very much for taking the time to complete this questionnaire. Your feedback is valued and very much appreciated.

APPENDIX 2

Interview protocol

Interview protocol for teachers of year 13 statistics

Thank you for responding to the questionnaire. This interview will be recorded in order to ensure accuracy but you may request the recording be turned off at any time during the interview.

1. Time Series was first introduced into the NZ Statistics curriculum in 1992. Do you have any thoughts on the merits or otherwise of introducing Time Series analysis to the school curriculum?
2. There have been a lot of changes in the Statistics curriculum in recent years. Discuss the change from AS 3.1 to AS 3.8. Was it harder or easier to manage than some of the other changes?
3. Did you undertake any professional development on the teaching of time series? What did you do? Was it beneficial? If answer is YES: Why was it beneficial? If answer is NO: Why was it not beneficial?
4. Where did you source datasets and contextual information for teaching purposes?
5. Technology has played a large part in curriculum changes in many disciplines. Discuss you and your school's approach to managing technological change in general and with respect to AS 3.1/ 3.8.
6. Describe your use of technology when teaching time series.
7. Describe your students' use of technology when learning about time series.
8. Describe how your department managed the introduction of the new Achievement Standard, AS 3.8. Is implementation left up to individual teachers or are lead teachers nominated. If you have a lead teacher describe their role?
9. Describe the internal assessment development process for time series at your school.
   [Prompt on choice of context, sourcing of datasets, sourcing of background information etc.]
10. Describe any changes in assessment and student response when comparing internal assessments produced for AS 3.1 and AS 3.8?
    [Prompt:- Changes in assessment – setting, marking, length, research component and structure. Changes in student responses to assessment – student engagement, quality of student work, level of student reasoning, quality of learning outcomes, level of student insight and understanding.]
11. Has the workload for students changed between AS 3.1 and AS 3.8? If yes, how?
12. Has the workload for teachers changed between AS 3.1 and AS 3.8? If yes, how?
13. In terms of difficulty, how does AS 3.8 compare to the other Statistics Achievement Standards?
14. If you could design a new Achievement Standard for the internal assessment of time series, what would it look like? Why?
15. If resources and time were not constraints on your teaching practices, how would you teach time series differently compared to your current teaching practices?
16. Communication of findings is an important aspect of the Statistical Enquiry Cycle. Discuss your approach to teaching this aspect and how this may have changed compared to AS 3.1?
17. If you used the *iNZight* time series module for time series analysis, describe the process.
    [Prompt: – in particular was it easy for you to use, was it easy for your students to use, were there any technological problems? What are the advantages of *iNZight*? What are the disadvantages of *iNZight*?]
18. Describe any changes in your approach to teaching time series under AS 3.8 compared to AS 3.1.
19. Which time series Achievement Standard, AS 3.1 or AS 3.8, do you prefer teaching and why?

20. If your students had a choice of AS 3.1 or AS 3.8 which time series Achievement Standard do you think your students would have preferred and why?
21. In your opinion what is the main difference for your students in the teaching and learning of time series when comparing AS 3.1 and AS 3.8?
22. Is there anything else that you would like to add?

APPENDIX 3

Teacher participant information sheet

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

DEPARTMENT OF STATISTICS

Faculty of Science

Level 3, Maths/Physics Building

38 Princes Street

Auckland, New Zealand

Telephone 64 9 373 7599

Facsimile 64 9 373 7018

http://www.stat.auckland.ac.nz

YEAR 13 STATISTICS TEACHER INFORMATION SHEET

To : **The Year 13 Statistics teacher, NZ Secondary School**

Title: **The impact of curriculum changes on the teaching and learning of time series analysis in Year 13**

Researcher: Rachel Passmore

My name is Rachel Passmore and I am a student and staff member at the University of Auckland. I am currently trying to complete a Master of Science degree in Mathematics Education. As part of my degree I am required to conduct research for my thesis. I wish to invite you to participate in my research investigating the impact of curriculum changes on the teaching and learning of time series analysis in Year 13.

Your school has been selected since your Mathematics department is known to have an on-going commitment to professional development in Mathematics Education. All Year 13 Statistics teachers in your school's department of Mathematics will be invited to take part.

In 2013, students were assessed under the new Achievement Standard in time series for the first time and anecdotally teachers have reported significant improvements in student outcomes and quality of work. The purpose of this research is to develop a rigorous framework against which student outcomes under the old and new regime can be assessed to discover if the anecdotal evidence can be confirmed or disputed by formal research.

I wish to collect data for the study in a number of ways. From freely available NZQA resources I will be gathering exemplars of student work completed for the old and new Achievement Standards. I will also, subject to your consent, survey you about your perceptions of the curriculum change. This will be done using a questionnaire which should not take more than 30 minutes to complete, at a time convenient to you. In order to gain a more in depth understanding of how this curriculum change has impacted on the learning of time series analysis, I would also like to interview one Year 13 Statistics teacher, at a mutually convenient time in the school Mathematics department's resource or meeting room for about an hour. This teacher will be randomly selected from those participating at your school. All this information will be collected within a four week period. I would like to audiotape the interview in order to ensure that it is accurately reported. All audiotaping will only be done with the consent of the teacher involved. If you are interviewed you may have the recording turned off at any time. A suitably qualified adult from outside the University will transcribe the audio file. Your name will not be on the audio file. You will be offered the opportunity to edit the transcript of your recording and a copy of the audio file.

If you agree to participate, all the information I obtain will be reported or published in such a way that the teachers and the school will not be identified. All the data will be kept securely for six years and then destroyed, as is usual for research at this university. Paper files will be shredded, computer and audio files will be deleted. You may request a summary of my thesis. Data arising from this research will be used in my Master's thesis but may also be used in presentations at professional development days and may be submitted in a paper to a reputable academic journal.

Please be assured that your Principal has given their assurance that your choice to participate or not in this research will not affect your employment or your relationship with your school.

Participation in this research is completely optional. You may withdraw your participation any time up to 30 September 2014 without giving a reason. If you are willing to participate in this research, please fill in and return the attached consent form. Thank you very much for your time and help in making this research possible. If you have any questions feel free to contact me at any time:

Rachel Passmore

Email: r.passmore@auckland.ac.nz

My supervisor is:
Assoc. Prof. Maxine Pfannkuch

The University of Auckland

Department of Statistics

Private Bag 92019

Auckland 1142

Phone: 923 8794

Email: m.pfannkuch@auckland.ac.nz

The Head of Department is:

Professor Chris Triggs

The University of Auckland

Department of Statistics

Private Bag 92019

Auckland 1142

Phone 923 8750

For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142. Telephone 09 373-7599 extn. 83711.

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE on 13 May 2014 for a period of three years. Reference Number 011642.

APPENDIX 4

Teacher consent form

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

DEPARTMENT OF STATISTICS

Faculty of Science

Level 3, Maths/Physics Building

38 Princes Street

Auckland, New Zealand

Telephone 64 9 373 7599

Facsimile 64 9 373 7018

http://www.stat.auckland.ac.nz

YEAR 13 STATISTICS TEACHER CONSENT FORM

THIS CONSENT FORM WILL BE HELD FOR A PERIOD OF SIX YEARS

Title: **The impact of curriculum changes on the teaching and learning of time series analysis in Year 13**

Researcher: ***Rachel Passmore***

I have been given and have understood an explanation of this research project. I have been able to ask questions and have them answered.

- I agree to participate in this research.
- I agree to allow my questionnaire results to be used as part of this research.
- I understand that I may be randomly selected for interview and that if I agree it may be audiotaped. I understand that even if I agree, I may choose to have the audio recorder turned off at any time.
- I understand that, if interviewed, I will be offered the opportunity to edit my transcript and a copy of the audio-file.
- I understand that my decision to participate or not in this research will not affect my employment or relationship with my school.
- I understand that neither my name nor that of my school will be identified in any reports.

- I understand that I will be involved in this research study for approximately 30 minutes and if I am interviewed for a further 60 minutes and that information will be collected over a four week period.
- I understand that all the data will be kept in a secure filing cabinet for six year at the Department of Statistics or Department of Mathematics at University of Auckland. After six years all paper files will be shredded, computer files deleted and audio-files destroyed.
- I understand that data will be used in Rachel Passmore's Master's thesis and may also be used in professional development presentations and used in a paper submitted to a reputable journal.
- I understand that I can withdraw my participation in this research project up until 30 September 2014 without giving a reason.
- I would like to request a summary of the thesis when it is completed. Please send summary to the following email address:…………………………………………….

Name of Year 13 Statistics Teacher

Signed:____ Date:_________________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE on 13 May 2014 for a period of three years. Reference Number 011642.