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.
Informally testing the fit of a probability distribution model

Anna-Marie Grace Fergusson
Abstract

Although research regarding teaching probability distributions from a modelling perspective is in its infancy, teachers in New Zealand have been required to teach new data-driven approaches alongside more established theory-driven approaches since 2013. A review of documents produced to support teaching probability distributions from a modelling perspective led to the identification of a potential problem with the current method taught to Year 13 students in New Zealand for informally testing the fit of a probability distribution model. Therefore, this study focused on how teachers informally test the fit of a probability distribution model.

Seventeen teachers participated in a task-based interview consisting of six tasks, one of which was a randomised experiment. The teachers’ responses were analysed using thematic analysis. The analysis produced categories of criteria for sketching distribution shape and for testing the fit of a probability distribution model. Structural analysis based on a statistical modelling framework proposed by the researcher was employed to identify key knowledge components used and connected by teachers when probability distribution modelling.

The findings of this study suggest that the main issues with informally testing the fit of a probability distribution model are the lack of criteria for considering (1) sample size and (2) the expected variation of sample proportions. Both criteria are difficult for teachers to take into account without using a formal test for significance. Teachers also struggled to separate the real world from the model world in their thinking about probability distribution modelling, appearing to conceive of only two components: model and data. The researcher’s new informal test for the fit of a probability distribution model incorporates sample size and the expected variation of sample proportions within her simulation-based modelling visualisation tool. The tool reinforces the separation and connection of the real world and the model world through the use of the researcher’s statistical modelling framework.
Acknowledgements

My love of teaching statistics and desire to learn more about what makes for excellent statistics education was cultivated during my time teaching at Avondale College. Therefore, I would like to acknowledge the support and encouragement I received from senior management and my department during the 12 years I worked at this school.

The opportunities I was given to lead mathematics and statistics within the school also led to my meeting with Pip Arnold. Working with Pip has positively impacted my knowledge of and involvement with curriculum design, assessment design and professional development for teachers.

I am very grateful for the expert guidance and advice I have received from my supervisor for this research project, Maxine Pfannkuch. If it was not for Maxine, I would not have moved into this new role as statistics education researcher. I see it as an unbelievable privilege that I am able to learn from Maxine and I am so appreciative of her patience and encouragement during this research journey.

I would also like to thank the fantastic members of the stage one statistics teaching team at the University of Auckland and our HOD Ilze Ziedins, who have all been so supportive of my study.

Lastly, I need to acknowledge the unwavering support of my husband Leo. He has helped me turn my ideas and thinking into a (hopefully) coherent piece of writing. Leo has looked after all the non-statistical aspects of our lives and this dissertation would not have been completed without him.
# Table of Contents

Abstract.................................................................................................................................................. 2
Acknowledgements................................................................................................................................. 3
List of Figures........................................................................................................................................... 7
List of Tables........................................................................................................................................... 9

Chapter 1 Introduction............................................................................................................................ 10
  1.1 Background..................................................................................................................................... 10
  1.2 Rationale for research....................................................................................................................... 11
  1.3 Research questions......................................................................................................................... 11
  1.4 Outline of dissertation chapters .................................................................................................... 12

Chapter 2 The problem........................................................................................................................... 13
  2.1 Introduction..................................................................................................................................... 13
  2.2 Review of government resources and assessments ......................................................................... 13
    2.2.1 Methods related to distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities........ 14
    2.2.2 Potential issues with teaching the method informally testing the fit of a probability distribution model ................................................................. 17
  2.3 Review of commercial and teacher-published resources.............................................................. 25
    2.3.1 Understanding the distribution of experimental estimates of probabilities.......................... 26
    2.3.2 Teaching the method within the wider probability modelling process ..................................... 30
    2.3.3 Teaching the method alongside statistical inference ............................................................... 33
  2.4 Conceptual map of potential issues................................................................................................. 40
  2.5 Development of research questions................................................................................................. 41

Chapter 3 Literature review .................................................................................................................. 42
  3.1 Introduction..................................................................................................................................... 42
  3.2 Tests for goodness of fit.................................................................................................................. 42
    3.2.1 Characterising an informal test for goodness of fit ................................................................. 42
    3.2.2 The chi-square goodness-of-fit test......................................................................................... 44
    3.2.3 Existing research for informally testing the fit of a probability distribution model................. 46
  3.3 Comparing distributional shape features ...................................................................................... 48
3.3.1 Existing research for sketching distributional shape .......... 48
3.3.2 Sketching distributional shape from a modelling perspective .......................................................................................... 50
3.3.3 The concept of overfitting models ........................................... 53
3.4 Towards a new informal test for the fit of a probability distribution model .............................................................................. 55
  3.4.1 Simulation-based visual inference ............................................. 56
  3.4.2 Proposed statistical modelling framework ................................. 59

Chapter 4 Methodology ................................................................................................................................. 64
  4.1 Introduction ................................................................................................. 64
  4.2 Mixed methods research ........................................................................... 64
  4.3 Use of interviews ...................................................................................... 66
  4.4 Qualitative data analysis .......................................................................... 68
  4.5 Research design ........................................................................................ 69
    4.5.1 Mixed methods design ...................................................................... 69
    4.5.2 Selection of participants ..................................................................... 70
    4.5.3 Data collection methods ..................................................................... 70
    4.5.4 Task-based interview design .............................................................. 71
    4.5.5 Analysis methods ................................................................................ 80
  4.6 Validity, reliability, ethics and researcher bias ........................................ 80

Chapter 5 Results ........................................................................................................................................... 84
  5.1 Introduction ................................................................................................ 84
  5.2 Task completion analysis ......................................................................... 84
  5.3 Results from Task B: Preparing to teach AS91586 ............................... 84
    5.3.1 Key understandings described by teachers ........................................ 84
    5.3.2 Understanding of the three types of probability distributions ................ 87
    5.3.3 Summary ............................................................................................ 88
  5.4 Results from Task C: Sketching shapes of distributions ......................... 89
    5.4.1 Similarity of the shape sketches drawn .............................................. 89
    5.4.2 Further categorising shape sketches .................................................. 92
    5.4.3 Summary ............................................................................................ 94
  5.5 Results from Task D: Assessing student understanding ......................... 95
    5.5.1 Describing the shape of small samples .............................................. 95
List of Figures

Figure 1: Reprinted from NZQA, 2015, AS91586 exam, Q2(b)................................................................. 16

Figure 2: Data displays and reasoning used reprinted from Lateness: Choice or chance? from the Senior Secondary Guide (Ministry of Education, 2013)........ 19

Figure 3: Reprinted from NZQA, 2015, Scholarship Statistics exam, Question 4b. ................................................................. 21

Figure 4: Shape sketch versus model overlay................................................................. 23

Figure 5: Key concepts reprinted from Probability distributions –What are the key ideas and how do we teach them (Martin, 2014). ................................................................. 28

Figure 6: “Post it note” diagram reprinted from Statistical stories and surprises: Probability distributions (Addison, 2015)................................................................. 29

Figure 7: Triangular distribution selection and justification diagram reprinted from Probability distributions –What are the key ideas and how do we teach them (Martin, 2014). ................................................................. 30

Figure 8: Sample data used by Barton and Laverty (2013, p. 480) with a Poisson model ($\lambda = 1.9$) overlaid ($\chi^2 = 14.12$, $p = 0.007$). ................................................................. 32

Figure 9: Probability distribution shape sketching activity reprinted from Probability distributions –What are the key ideas and how do we teach them (Martin, 2014). ................................................................. 34

Figure 10: Modelling process diagram reprinted from Probability distributions –What are the key ideas and how do we teach them (Martin, 2014). ................................................................. 35

Figure 11: Screenshot of spreadsheet tool designed by Addison (2015).......................... 39

Figure 12: Conceptual map of knowledge required for AS91586 ........................................ 40

Figure 13: An example of an over-fitted distributional shape sketch...................... 49

Figure 14: Two examples of sketching shape for the same sample distribution. 51

Figure 15: Screenshot of VIT randomisation module to test the difference between two medians.................................................................................................................. 57

Figure 16: Graphical inference applied to testing a probability distribution model, (m) is the real sample data distribution. All other plots are simulated samples from a discrete uniform distribution model (a fair six-sided die) .......... 58

Figure 17: Three interconnected ways to think about probability, reprinted from Pfannkuch and Ziedins (2014, p. 104). ................................................................. 60

Figure 18: Key components of the proposed statistical modelling framework ... 62
Figure 19: Diagram of procedures for this study ............................................................... 69
Figure 20: The online environment used for the task-based interview ......................... 71
Figure 21: Sample distribution used for Task D ............................................................... 74
Figure 22: Key design features of Task E ........................................................................ 76
Figure 23: Screenshots demonstrating the use of the tool to informally test the fit of a probability distribution model ................................................................. 79
Figure 24: Screenshot of guided exploration used in Task F ........................................... 80
Figure 25: Task C shape sketches for each distribution produced by combining the sketches from all 12 teachers and overlaid on the same plot without the reference distribution ......................................................................................... 89
Figure 26: Example of sketches of the sample distribution used in Task D .............. 96
Figure 27: Framework analysis using an adaption of Martin's (2016) statistical modelling framework that includes knowledge components ................................................. 103
Figure 28: The two versions of Task E, presented with examples of good responses ...................................................................................................................... 108
Figure 29: Version 2 of Task E, which shows the sample distribution (n = 480) with the model probability distribution overlaid on the same graph ............ 114
Figure 30: An example of how the new simulation-based modelling tool could have been used by Teacher 13 to test the fit of the probability distribution model (see section 4.5.4 for more information) ......................................................... 115
Figure 31: The two versions of Task E presented with screenshots of visualisations of sampling variation from Wild et al. (2011) with similar sample sizes ............................................................................................................................. 119
Figure 32: Combining model-driven and data-driven approaches in the proposed statistical modelling framework ................................................................. 123
Figure 33: An example of a distributional shape sketch using a thick line .......... 127
Figure 34: Example of thinking about a problem using the statistical modelling framework. Problem based on NZQA AS91585 2015 Q2(a) .................. 129
Figure 35: NZQA AS91586 2015 Q2(b) model fit test using new simulation-based modelling tool .................................................................................................................. 130
List of Tables

Table 1: Commercially published documents used in review........................................ 25
Table 2: Census At School (NZ) published documents used in review ....................... 25
Table 3: Terms used when discussing the distribution of true probabilities, from a review of the documents written by Hinchliffe and Priest (2013) and Lakeland and Nugent (2016).......................................................................................................................... 27
Table 4: Example of calculations to support chi-square goodness-of-fit test ............. 45
Table 5: Summary of tasks used within interview.......................................................... 71
Table 6: The six distributions used in Task C, cross-classified by sample size and potential modality .................................................................................................................. 73
Table 7: Key understandings for learning AS91586 described by teachers .............. 86
Table 8: Comparison of shape sketches produced for Task C ...................................... 91
Table 9: Arnold’s (2013) criteria for sketching the shape of the distribution of a variable........................................................................................................................................ 92
Table 10: Criteria for sketching the shape of the distribution of a variable for general characteristics.................................................................................................................. 94
Table 11: Shape sketches for Task E, matched by pairs of visually similar sketches across Versions 1 and 2 of the task .....................................................................................106
Table 12: Categories of criteria described to sketch the shape of a distribution .................................................................................................................................107
Table 13: Categories of criteria used to assess the visual fit of the model.............. 110
Table 14: Number of students assessed in 2015 for various curriculum level eight achievement standards ........................................................................................................... 118
Table 15: Terms used when discussing each type of probability distribution, from a review of the documents written by Hinchliffe and Priest (2013) and Lakeland and Nugent (2016).......................................................................................................................... 156
Chapter 1

Introduction

“Statistical thinking and probabilistic thinking are two sides of the same coin in the sense that practitioners faced with a problem will draw on either one or both perspectives and methods….. The interaction and interconnection of statistical and probabilistic knowledge and thinking present a problem for novices who need to determine when to use one or the other or both” (Pfannkuch et al., 2016, p. 13).

1.1 Background

For five years I held the position of Director of Mathematics and Statistics at a very large co-educational secondary school in Auckland, New Zealand. In this role, I was responsible for leading the implementation of the New Zealand Curriculum (Ministry of Education, 2007) and overseeing the development of assessments for the re-aligned National Certificate of Educational Achievement (NCEA) achievement standards. At a national level, I was also a member of the standards writing group for the NCEA Level 2 and 3 Statistics achievement standards and a member of other groups charged with developing a wide range of associated learning and assessment material for the Senior Secondary Guide for Mathematics and Statistics (Ministry of Education, 2013) and the New Zealand Qualifications Authority.

Through my curriculum and assessment development work, both within my department and also within national projects, I become very aware of the varied ways data and simulations are used to teach statistics and probability, and the potential for confusions of key ideas and concepts as a result. In the area of probability distribution modelling, I was particularly interested in conducting research with teachers to explore how data is used within an area of teaching that has been dominated by theoretical approaches.
1.2 Rationale for research

The implementation of the New Zealand curriculum (Ministry of Education, 2007) had a significant impact on the nature of the statistics taught and assessed in secondary schools. New methods were introduced into the curriculum, such as the use of re-sampling or bootstrap confidence intervals and randomisation tests, and the need for students to connect data and probability was more clearly articulated as an expectation for learning.

AS91586 Apply probability distribution in solving problems was a new achievement standard created through the re-alignment process. AS91586 requires students to demonstrate understanding of methods related to the distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities. Students need to compare and reason informally about empirical distributions and theoretical probability distributions. However, research associated with teaching and learning probability modelling is still new and as such there is not yet an established common body of knowledge for teachers to use when engaging with students in the probability modelling process. Therefore, both within the unique New Zealand curriculum requirements and in the international perspective, there was a need for further research into the area of teacher knowledge of probability modelling involving empirical data.

1.3 Research questions

This study explored teachers’ knowledge of probability distribution modelling as related to the curriculum achievement objectives associated with AS91586 Apply probability distribution in solving problems. Because a significant change to the curriculum required students to compare model probability distributions with empirical distributions, this study focused on teachers’ knowledge of informally testing the fit of a probability distribution model.
The main research question for this study is:

How do teachers informally test the fit of a probability distribution model?

Supporting research questions are:

1. How do teachers understand "the distribution of experimental estimates of probabilities"?

2. How is the informal test of the fit of a probability distribution model taught within the wider probability distribution modelling process?

3. How are statistical inference concepts used when teaching probability distribution modelling?

1.4 Outline of dissertation chapters

Chapter One provides the researcher's background and interest in the research, followed by discussion of the need for the research and the presentation of research questions. In Chapter Two, a review is conducted of government and teacher-produced documents for supporting the teaching of curriculum achievement objectives assessed by AS91586. The current informal test for the fit of a probability distribution model is described and the potential issues associated with the test are discussed. Chapter Three presents a review of literature related to teaching an informal test for the fit of a probability distribution model. The literature is used to support the development of a new informal test for the fit of a probability distribution model and a new proposed statistical modelling framework. In Chapter Four, the research methods and design used for the study are presented, including discussion on the nature of the tasks designed and used for the research. Chapter Five presents the results from the analysis of each of the six tasks. Chapter Six concludes the dissertation with a discussion on answers to the research questions, followed by consideration of the limitations of the study, related areas for further research and teaching and assessment implications.
Chapter 2
The problem

2.1 Introduction

In New Zealand, students undertaking the National Certificate in Educational Achievement (NCEA) at Level Three (Year 13) can be assessed against seven different achievement standards based on curriculum level eight achievement objectives from the statistics strand of the New Zealand Curriculum (Ministry of Education, 2007). AS91586 *Apply probability distributions in solving problems* is one of these NCEA Level 3 statistics Achievement Standards and is externally assessed by the New Zealand Qualifications Authority (NZQA) through a one hour written exam paper. Students who are assessed against AS91586 *Apply probability distributions in solving problems* are required to investigate situations that involve elements of chance using methods such as calculating and interpreting expected values and standard deviations of discrete random variables, and applying discrete and continuous probability distributions (e.g. uniform, triangular, Poisson, binomial, and normal). Students are also required to use methods related to *distribution of true probabilities* versus *distribution of model estimates of probabilities* versus *distribution of experimental estimates of probabilities*.

This chapter explores the meaning of terms *distribution of true probabilities*, *distribution of model estimates of probabilities*, and *distribution of experimental estimates of probabilities* and discusses potential issues with the assessment of methods related to these terms, in particular the problem of teaching a method for *informally testing the fit of a probability distribution model*. This discussion includes a review of the resources available within New Zealand to support the teaching of curriculum achievement objectives associated with AS91586.

2.2 Review of government resources and assessments

In this section, government produced curriculum guides and assessment items in relation to AS91586 are explored and analysed resulting in potential issues being
raised about teaching the method *informally testing the fit of a probability distribution model*.

### 2.2.1 Methods related to distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities

To support teaching of levels six through eight (Years 11 to 13) of the New Zealand Curriculum, the Ministry of Education created an online resource called the Senior Secondary Guide¹ (Ministry of Education, 2013). Within the *Understanding true probability, model estimates, and experimental estimates* teacher notes on the Senior Secondary Guide², there are no specific explanations or examples given for the *distribution of true probabilities*, the *distribution of model estimates of probabilities* or the *distribution of experimental estimates of probabilities*. Instead the teacher notes state that “... we can also think in the same three ways about the probabilities of a collection of events or a probability distribution of outcomes but these are not addressed in the discussion ...”. An excerpt from these teacher notes states:

> The model probability or theoretical probability of an event is the probability assigned under a given model. The experimental probability of an event is the probability obtained from trials or simulations, which are based on some underlying assumptions (for example, independence of trials). Both the model probability and the experimental probability of an event give estimates of the “true” probability. Both of these methods for determining the probability of an event are interconnected and they both are seeking to determine the “true” probability of an event, which is usually unknown.

Also within these teacher notes is an explanation that “...experimental data in probability can be any results of observation of the situation...” and that experimental data can be used to calculate the experimental probability. The

---


teacher notes also explain that “...the experimental estimate of probability can be compared with the model estimate to evaluate whether the model is an accurate representation of the context...”.

An interpretation of these teaching notes is that a method, which is related to all three different ways to think about probability distributions, is *informally testing the fit of the probability distribution model*. The method would involve comparing the model probability distribution with the experimental probability distribution, and concluding that if the model was a “good fit” then the model is a “good model” for the true probability distribution. This interpretation is consistent with the assessment items that have been developed by NZQA for external assessment of AS91586, an example of which is the assessment item shown in Figure 1a.

In this assessment item, the three “types” of probability distributions are:

- The true probability distribution: The number of bus breakdowns per 12-hour period, where both the sample space (set of outcomes) and likelihood of each outcome are unknown
- The model probability distribution: A Poisson distribution model with $\lambda = 3.6$
- The experimental probability distribution: A sample of 12-hour periods collected over a “long period of time”, displayed in terms of observed outcomes and their relative frequencies (proportions)

Figure 1b shows the expected response for this assessment item, and demonstrates the method *informally testing the fit of the probability distribution model*, which is as follows: compare the features of the model and experimental probability distributions and then use these comparisons to judge the “goodness” of the model as it applies to the true probability distribution. However, there are potential issues with teaching the method *informally testing the fit of the probability distribution model* when the learning of probability distribution modelling is placed within the wider context of the learning of statistics at curriculum level eight (Year 13) of the New Zealand Curriculum.
(a) assessment item

Image removed from final copy due to copyright.

(b) expected response

Image removed from final copy due to copyright.

Figure 1: Reprinted from NZQA, 2015, AS91586 exam, Q2(b).
2.2.2 Potential issues with teaching the method informally testing the fit of a probability distribution model

From the researcher’s analysis of the government resources regarding the teaching of informally testing the fit of a probability distribution model, there are three main potential issues. These issues, which are now discussed, are: using the word “experimental”; teaching the method within the wider probability modelling process; and teaching the method alongside statistical inference.

Using the word “experimental”

According to the Glossary of statistics terms available on the Senior Secondary Guide\(^3\), the simplest meaning of the word experiment is “... a process or study that results in the collection of data, the outcome of which is unknown”. The glossary also states that an experimental distribution can also be called an empirical distribution and that a sample distribution is sometimes called an experimental distribution.

However, in the long history of teaching theoretical and experimental probabilities in New Zealand schools, the term experimental distribution is often associated only with probability experiments. Consequently, the associated term experimental estimate of probability has been taken to mean as it has been defined in the glossary, as “an estimate of the probability that an event will occur calculated from trials of a probability activity by dividing the number of times the event occurred by the total number of trials”.

While the meaning of experimental probability is necessarily expanded within the Understanding true probability, model estimates, and experimental estimates teacher notes to include not just trials of a probability activity but also the results of “...simulations, or the results of observation of the situation ...”, the historic use of the term experimental probability is not the only issue. When using data from sampling, probability activities and simulations to create a distribution of

experimental estimates, there is a danger that the understanding of what this distribution is an estimate of, may not be clearly understood.

The Understanding true probability, model estimates, and experimental estimates teacher notes, for example, state that “...both the model probability and the experimental probability of an event give estimates of the “true” probability...”, however, this is not the case when using an experimental distribution generated from a simulation. Because a simulation requires a model to be formed, any experimental distribution of simulation results is in fact a distribution of experimental estimates for the model probabilities. Therefore, data collected through sampling or through observations of the situation (the situation being one where the distribution of true probabilities is unknown) needs to be viewed as different from data generated from a model through simulation, which may be an issue if both are described as experimental distributions.

The teaching notes treat both sample data and data generated through simulation as equivalent in terms of them being the distribution of experimental estimates of probabilities. However, the Lateness: Choice or chance? example⁴, also on the Senior Secondary Guide, clearly separates data collected through sampling from the actual situation being modelled, from data generated from the model being considered for the situation. The data collected through sampling (called raw data) is then compared to the data simulated from the probability distribution model to see if they are similar. The Lateness: Choice or chance? example uses bar charts of each distribution (sample versus simulated) with the model distribution overlaid as a line graph to assist visual judgements as to model fit (see Figure 2). While this approach makes sense, there is still an issue concerning how a student judges whether the sample distribution is similar to the simulated data from the model probability distribution based on visual appearances.

---
(a) Data collected from situation with probability distribution model
distribution overlaid

Image removed from final copy due to copyright.

(b) Simulated data from the probability distribution model for the situation
with probability distribution model overlaid

Image removed from final copy due to copyright.

(c) Reasoning used to test fit of probability distribution model
Students should be able to identify that while the theoretical model and simulated data (for
200 lessons) do not match perfectly, the mode of “no students” from the raw data is not
occurring in the simulated data, so it does not appear that the binomial model (with n = 30 and
p = 0.045) is a good model, even when taking into account the variation expected from having
data from only 200 lessons.

Figure 2: Data displays and reasoning used reprinted from Lateness: Choice or
Teaching the method within the wider probability modelling process

There is also a potential issue with teaching the method *informally testing the fit of the probability distribution model* within the wider modelling process. Two specific issues are how to integrate knowledge from different sources when modelling and how to consider the purpose of the model.

**Integrating different knowledge sources when modelling.** The expected response shown in Figure 1b includes discussion on the appropriateness of the model in terms of assumptions of the Poisson distribution. This discussion uses knowledge about the context (buses on the road, day versus night) and knowledge about the probability distribution model being tested (need for constant rate for Poisson distribution). The expected response also shows use of the method *informally testing the fit of the probability distribution model*, which could be described as combining knowledge about the probability distribution model with knowledge about the sample data.

The potential issue comes from how to teach students to integrate these different knowledge sources when modelling. The expected response in Figure 1b is not that helpful in this respect, as both parts of the response come to the same conclusion; that the Poisson distribution is not a good model for the number of buses that breakdown in any 12-hour period. What is not demonstrated clearly in the related materials in the Senior Secondary Guide (Ministry of Education, 2013) nor the assessment items developed by NZQA is how students should reach a decision about whether a probability distribution model is a good model if *informally testing the fit of the probability distribution model* leads to a conclusion that is not the same as the decision reached when comparing contextual and probability distribution model knowledge.

**Considering the purpose of the model.** An example of an assessment item that assesses the method *informally testing the fit of the probability distribution model* in isolation from the wider probability modelling process is shown in Figure 3. This assessment item is from the NZQA Scholarship Statistics, an exam which assesses the same curriculum achievement objectives as AS91586. The data presented are historic sales data for a real estate office, and the assessment item
requires students to test the fit of two different probability distributions models to the data.

However, students are only asked to estimate a probability for an event based on randomly selecting one of the houses in the supplied data. Therefore, there is no real purpose in considering or testing a probability distribution model, as within this context the distribution of experimental estimates of probabilities is exactly the same as the distribution of true probabilities. That is, if a probability distribution model was to be defined, it should be exactly the same as the distribution of experimental estimates of probabilities in order to best answer the question.

Furthermore, there is no need to generalise the features of the sample data provided to make an inference about the underlying population distribution, and the assessment item is only concerned with the 325 houses that the real estate office sold in the past year. The only mathematical issue here is that the required probability does not correspond to the class intervals of the histogram. There is
no purpose in requiring a probability distribution model because the data is available and the histogram could be easily re-drawn.

Additionally, even if the assessment item had instead asked for the students to use each model to estimate the probability that the next house sold by the real estate office sold for between $450,000 and $750,000, the use of the method *informally testing the fit of the probability distribution model* would still be carried out in isolation from the probability modelling process. This is because earlier in the same exam paper, information had been presented that demonstrated that selling prices of houses had been changing over the past several years. Therefore, it would be difficult to justify that the features of the sales data for one year could be generalised and used to model the features of sale data for future years in way that would be useful. That is, while the probability distribution model might be a good fit for the sample data, it might not be a good model for the true probability distribution.

**Teaching the method alongside statistical inference**

Both of the assessment items shown in Figure 1a and Figure 3 use sample data collected from the situation being modelled. Consequently, the measures or features of the sample data are only estimates of these same measures or features of the underlying population from which the sample has been taken, and the accuracy of these estimates is affected by the size of the sample, as well as the quality of the sample. This is knowledge that students are expected to use when making statistical inferences.

Students participating in learning programmes that teach statistical inference methods (e.g. confidence intervals, randomisation test) alongside the method *informally testing the fit of the probability distribution model* may be confused by the similarities and differences between the methods. In particular, issues may arise when the nature of the data and the amount of data is not considered when comparing measures of probability distribution models and sample distributions, specifically when sketching the shape of the sample distribution, and when making a call about the “goodness of fit”. 
**Sketching the shape of a sample distribution.** While the physical sketch of the shape of a distribution has not yet been required for the assessment of AS91586, describing the shape of a sample (experimental) distribution is included in the expected response shown in Figure 1b. Therefore it would make sense that students have experienced sketching the shape when learning about probability distribution modelling. Students also sketch and describe the shape of a sample distribution when learning about statistical inference. However, in the statistical inference learning context rather than compare the shape of the sample distribution to the shape of a known model distribution, students are required to infer the shape of the underlying population distribution, which is unknown.

The potential issue is whether the sketching of the shape of a sample distribution means the same thing when learning about statistical inference and probability distribution modelling. In both situations, it is not clear how the amount of data in the sample distribution should influence the sketching of its shape, including how much “smoothing” should take place when sketching the shape. Additionally, there is visual similarity between a shape sketch and overlaying of a model distribution on a sample distribution (see Figure 4).

![Figure 4: Shape sketch versus model overlay](image)

One difference between a shape sketch and a theoretical model overlay, observed by the researcher in student work, is that students tend to produce shape sketches that go above the distribution (see Figure 4a), while model overlays tend to go through the distribution due to density principles (see Figure 4b).
Making a call about the “goodness of fit”. Associated with the possible confusion about learning statistical inference methods alongside the method informally testing the fit of the probability distribution model is that there are no clear criteria for making a call about the “goodness of fit”. This is in contrast to how students are expected to reason with data when making statistical inferences. For example, students who are assessed against AS91583 Conduct an experiment to investigate a situation using experimental design principles (another NCEA Level 3 achievement standard) are required to use the randomisation test to assess the strength of evidence against the chance acting alone model. This informal procedure uses simulation to create a re-randomisation distribution which students then use to compare what they observed (e.g. the result of their experiment) to what they would expect under chance acting alone (what the result could look like if chance was acting alone).

While the New Zealand Curriculum (Ministry of Education, 2007) expects that students learning about probability “…acknowledge samples vary…”, “…compare and describe the variation between theoretical and experimental distributions…” and appreciate “…the role of sample size…” (excerpts from the probability thread, curriculum levels three to six), there is no guidance provided as to how to do this objectively. Instead, subjective evaluations are made when comparing theoretical and experimental distributions, using words such as “close” or “similar”. In fact, the expected response shown in Figure 1b does not even support the conclusion that the “…Poisson model is not a good model for the distribution of bus breakdowns for any 12-hour period…” with any specific points of comparison. Furthermore, the assessment item shown in Figure 1b does not specify how much data has been collected from the situation being modelled, only that the number of breakdowns has been monitored over “…a long period of time.”
2.3 Review of commercial and teacher-published resources

Various material has been developed by New Zealand teachers in response to the requirements of teaching the curriculum achievement objectives that are assessed by AS91586. The researcher attempted to source a range of workbooks, textbooks or other commercially published material that had been written to support teaching and learning Achievement Standard 91586. The researcher also accessed the Census At School New Zealand website and reviewed the material listed under teaching resources associated with AS91586. Documents were selected that were comprehensive and included details of the author. Hence, three documents published on the Census At School New Zealand website and four commercially published documents were reviewed in terms of potential issues with teaching the method *informally testing the fit of a probability distribution model* when the learning of probability distribution modelling (see Tables 1 & 2).

### Table 1: Commercially published documents used in review

<table>
<thead>
<tr>
<th>Type of document</th>
<th>Title</th>
<th>Author and date of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workbook</td>
<td>NCEA Level 3 Statistics Learning Workbook</td>
<td>Hinchliffe &amp; Priest, 2013</td>
</tr>
<tr>
<td>Workbook (online)</td>
<td>Sigma Statistics NCEA Level 3</td>
<td>Barton &amp; Laverty, 2013</td>
</tr>
</tbody>
</table>

### Table 2: Census At School (NZ) published documents used in review

<table>
<thead>
<tr>
<th>Title of document</th>
<th>Author and date of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of true probabilities and other new ideas&lt;sup&gt;5&lt;/sup&gt;</td>
<td>Steel, 2013</td>
</tr>
<tr>
<td>Probability distributions – What are the big ideas and how do we teach them?&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Martin, 2014</td>
</tr>
<tr>
<td>Statistical stories and surprises: Probability distributions&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Addison, 2015</td>
</tr>
</tbody>
</table>

The potential issues already identified in section 2.2.2 were present in the documents reviewed. However, in this review of commercial and teacher-published resources further insight was gained. The additional potential issues are discussed using similar themes to section 2.2.2, which were: using the word “experimental”; teaching the method *informally testing the fit of a probability distribution model* within the wider probability modelling process; and teaching the method *informally testing the fit of a probability distribution model* alongside statistical inference.

### 2.3.1 Understanding the distribution of experimental estimates of probabilities

Six of the seven documents reviewed included at least one definition, explanation or example concerning *true probability* versus *model estimate of probability* versus *experimental estimate of probability*. The use of the word “experimental” appears to be an issue across these documents, as it was in the curriculum and assessment material, compounded by the amount of vocabulary used when explaining the different ways of thinking about probability distribution and by the use of diagrams.

**Amount of vocabulary used.** Unlike the Senior Secondary Guide teacher notes, two of the commercially published documents specifically discussed *distribution of true probabilities* versus *distribution of model estimates of probabilities* versus *distribution of experimental estimates of probabilities* (Hinchliffe & Priest, 2013; Lakeland & Nugent, 2016). The following is an excerpt from the explanation given by Lakeland and Nugent:

> True data is the real life distribution of data. This data is usually unknown to the researcher. It could, for example, be the expected life of computer hard drives. Assuming you cannot run every hard drive until it fails, the researchers will never know the actual distribution. Let us assume the real life or true probability distribution is as per this diagram. (p. 3)

Reviewing the language used across these two documents uncovered many examples of different terms being used to describe the same concept. An example
of the variety of terms used when discussing the distribution of true probabilities in shown in Table 3 (see Appendix C for the full table). Adjectives vary in use between true, population, real, real world, real life, actual and unknown, and nouns used include some that are not statistically equivalent, for example, data, distribution and probabilities.

Table 3: Terms used when discussing the distribution of true probabilities, from a review of the documents written by Hinchliffe and Priest (2013) and Lakeland and Nugent (2016)

Terms related to a distribution: True probabilities, Population probability distribution, Population distribution, True distribution of probabilities, Actual distribution, Real life probability distribution, True probability distribution, True distribution, Real distribution, Distribution, True data distribution, Data usually unknown, True data, Real life distribution of data

Terms not specifically related to a distribution: Usually unknown, Unique to situation, What would actually happen in real world, Actual situation, Situation, True situation, Real situation

The word “situation” is also used in way that suggests this is not the same as the true probability distribution, and the use of the word population does not clearly signal that the situation being explored must involve a chance element for there to be a corresponding true probability distribution.

The interchangeable use of the terms data, distribution and probabilities continues when the distribution of model estimates of probabilities and the distribution of experimental estimates of probabilities are discussed. In fact, within the explanations concerning distribution of experimental estimates of probabilities, further terms are introduced, such as simulation results, sample data, long-run relative frequencies and graph. All these terms make the language involved complex and potentially confusing. For the remainder of this section, the researcher has attempted to use consistent language when describing the
guidance provided in documents for clarification purposes, which was not a feature of the language used in the documents.

**Visual diagrams for incorporating the distribution of experimental estimates of probabilities.** In addition to the potential issue with understanding *distribution of experimental estimates of probabilities* due to the variety, amount and conflicting nature of the vocabulary used, a new issue was identified by reviewing two of the documents published on Census At School New Zealand. The issue was the representation of the relationship between *the distribution of experimental estimates of probabilities* and the true and model probability distributions.

The diagram developed by Martin (2014) does not make reference to the notion of *true probability* or a *true probability distribution*, and instead refers to a “Random variable” (see Figure 5).

![Figure 5: Key concepts reprinted from Probability distributions – What are the key ideas and how do we teach them (Martin, 2014). Reprinted with permission.](image)

Similar to the teacher notes from the Senior Secondary Guide, the word “experimental” is used to describe both simulated and sample data. However, the use of arrows on the diagram suggests that both sample and simulated data can be collected directly from the random variable, and also both sample and simulated data can be generated from the model probability. While this diagram demonstrates that experimental data can be used as part of the modelling process
alongside or to support the development of theoretical probability distribution models, it does not keep the dual roles of data separate.

Similarly, in a “post it note” diagram developed by Addison (2015), the experimental estimate of the true probability is able to be determined using both real life data and from a simulation, keeping sample data and simulated data together visually (see Figure 6).

![Figure 6: "Post it note" diagram reprinted from Statistical stories and surprises: Probability distributions (Addison, 2015).](image)

What is not shown in either of these diagrams is that simulated data is only generated from a model. Both diagrams with the use of arrows give the impression that simulated data can be generated without a model and used directly to estimate the true probability.
2.3.2 Teaching the method within the wider probability modelling process

Five of the seven documents reviewed gave greater insight into how the method informally testing the fit of a probability distribution model might sit within the wider probability modelling process. However, there were still potential issues identified with how to integrate contextual knowledge and knowledge of the sample distribution, and also with some examples of sample data being used in the probability modelling process with no method used to check the fit of the probability distribution model to the sample data.

**Integrating contextual knowledge and sample distribution knowledge.** A strong similarity between the three documents published by Census At School (NZ) was the need for both contextual knowledge and knowledge of the sample distribution (features etc.) to be used in deciding on the appropriateness of a probability distribution model. However, how the two sources of knowledge were to be considered was still not made clear.

When explaining probability distribution model selection and justification Martin (2014) includes the diagram shown in Figure 7. This diagram demonstrates that both the features of the sample distribution (specifically its shape) and features of the context should match the features of the model probability distribution, but the specifics of how to conclude whether the model overall is appropriate or a good fit are not given.

![Triangular distribution selection and justification diagram](image)

*Figure 7: Triangular distribution selection and justification diagram reprinted from Probability distributions –What are the key ideas and how do we teach them* (Martin, 2014). Reprinted with permission.
Steel (2013) also discusses the possible selection of a triangular distribution as a probability distribution model, and states that this model can be used if “... the context and shape of the sample distribution suggest a triangular distribution”. This could imply that both the features of the sample distribution (specifically its shape) and features of the context should match the features of the model probability distribution in order to accept this distribution as a “good” model. However, later in her document, Steel states that “…whether a sample distribution is consistent with being from a population which could be modelled by a given distribution is a matter of judgement...” and that whether the model is useful (based on contextual knowledge) should also be a consideration.

**Sample data but method not used.** Two of the four commercially published documents did not provide guidance on how to informally test the fit of a probability distribution model. However, both included at least one question that required a probability distribution model to be fitted to sample data in order to answer it. Both documents presented the sample data in tabular form and neither document required the fit of the Poisson model to be tested instead only requiring that the sample mean was used as an estimate for the value of lambda for the Poisson distribution. Therefore, sample data were used in the probability distribution modelling process without an informal test for the fit of the selected probability distribution model to those data.

The potential issue with only using the data to obtain model parameters can be seen in the documents produced by Barton and Laverty (2013) and Hinchcliffe and Priest (2013). Both documents use data for the number of wasp nests found per defined square area and stated that a Poisson model can be used to model this situation. Figure 8 shows the sample data used by Barton and Laverty with the Poisson distribution model overlaid. Visually, there is a clear mismatch between the sample distribution and the model probability distribution. Additionally, a *Chi square test* gave very strong evidence against the Poisson distribution model. This was also the case for the sample data used for a very similar wasp nest question by Hinchcliffe and Priest.
Adopting a modelling perspective. All of the workbooks and textbooks reviewed gave explanations related to distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities. However, it was not clear how the questions that followed these explanations assessed these three types of probability distributions. Instead, the questions focused predominantly on theory-driven modelling. Even when sample or simulated data were expected to be used as part of the modelling process, the notion of a distribution of true probabilities, or more directly the notion of calculations from models as being estimates, was not assessed knowledge.

Two documents did, however, clearly articulate some characteristics of thinking consistent with a probability modelling perspective (Nugent & Lakeland, 2016; Steel, 2013). These characteristics can be summarised as:

- considering how well the features of the probability distribution model match the situation being modelled;
- considering how likely the data collected represents the true probability distribution;
- considering how much data was collected as increasing the amount of data should improve the reliability of the model parameters;
- considering the distribution of the data collected;
- considering the limitations of the probability distribution model used;
• considering the purpose of the probability distribution model and the level of accuracy required for that purpose.

On the one hand, these characteristics comprise a good list of factors to take into account alongside an informal test of the fit of a probability distribution model. On the other hand, a potential issue is about how teachers can integrate all of these understandings and concepts when working with students if the situation being investigated is one where contextual knowledge and the purpose of the model is limited.

2.3.3 Teaching the method alongside statistical inference

Six of the seven documents reviewed demonstrate the use of methods that could also be used within statistical inference learning contexts. Similar potential issues were identified about how to take into account sample size in regard to comparison of features such as measures of central tendency, spread and shape. An additional potential issue was identified regarding the use of simulation when informally testing the fit of a probability distribution model.

Sketching the shape of a probability distribution. It was noted earlier in Section 2.2.2 that while descriptions of shape for both the sample (experimental) distribution and the probability distribution model were used as part of the method informally testing the fit of a probability distribution model, there were no guidelines about how to sketch the shape of a sample (experimental) distribution. This was also the case in the documents reviewed.

Although Martin (2014) developed resources that prompt students to sketch probability distributions based on key features such as centre, spread, shape and sample space (see Figure 9), these sketches are done without reference to an existing distribution and would likely be sketched using features of theoretical probability distributions such as the triangular, uniform or Normal distributions. In this way, the students may consider these sketches representative of the probability distribution model.
In fact, elsewhere in Martin's document, when shape sketches are mentioned, they are in reference to the theoretical probability distribution model that has been selected (see Figure 10). Both examples shown in Figure 10 use sample distributions with the probability distribution model overlaid and it is not obvious by looking at the examples without reading the accompanying text whether the shape sketched is for the sample distributions or for the probability distribution model. Additionally, both sets of data appear to be large: 200 spins for Example 1; and millions for “the ages of married people on Facebook”. However, how the amount of data would affect these sketches or descriptions of shape is not discussed.

Two of the documents reviewed do specifically mention sample size in terms of specific numbers when discussing features of the sample distribution, including shape. Both Hinchcliffe and Priest (2013) and Steel (2013) suggest that sample distributions from large samples of around 200 should look similar to the population that they came from. However, the word “similar” is subjective, and there is still the issue of how to make a call about whether the model is “good” or not. In this respect, Steel (2013) is of the opinion that “… for assessment, sample distributions should either be reasonably obvious or the [assessment] schedule should accept a justified decision either way...”. However, “obvious” is still a
subjective word, and what is an “obvious” difference to one person may not be judged the same by another.

![Figure 10: Modelling process diagram reprinted from Probability distributions – What are the key ideas and how do we teach them (Martin, 2014). Reprinted with permission.]

**Making the call?** In addition to the examples provided by Martin (2014) as shown in Figure 10, two of the four commercially published documents provided examples that demonstrate the method *informally testing the fit of a probability distribution model*. Both of these documents compare features of the model probability distribution with features of the experimental probability distribution, and conclude that if the model is a “good fit” then the model is a “good model” for the true probability distribution. In examples that test the fit of a Poisson distribution model, specific steps for the method were:
• Calculate the mean and variance for the sample data. If these are similar, then the Poisson distribution is a good model.

• Compare individual model probabilities with the sample proportions. If these are similar, then the Poisson distribution is a good model.

• Identify and describe individual outcomes for the sample data with proportions that do not match the model probabilities.

However, neither document included guidance for how to take into account sample size when assessing the fit of the Poisson distribution model, nor gave criteria for how to decide if the features compared were similar. When considering the Poisson distribution as a model for the number of cars in a drive-through lane every five minutes, Lakeland and Nugent (2016) stated:

\[
\text{Mean} = 2.425 \ldots \text{ Variance} = 2.244 \ldots \text{ As variance is similar to mean it confirms the best model is Poisson.} \quad \text{(p. 58)}
\]

However, assessing how similar these are cannot be done without considering the sample size, a factor Lakeland and Nugent (2016) referred to earlier in their document “... increasing the data should improve reliability of the model parameters ...” (p. 4).

Additionally, neither document specifically discussed comparing the similarity of shape of the sample distribution and the model distribution. When considering the Poisson distribution as a model for the number of complaints about incorrect orders a company receives every week, Growing Minds (2016) stated:

\ldots as can be seen from the graph (with model estimates shown in blue) the model is a reasonable fit for experimental data. \ldots \quad \text{(section 25.2)}

The explanation has been taken by the researcher as a direction to students to compare the individual model probabilities with the sample proportions visually on the graph, rather than an overall comparison of the shapes of the sample and model distributions. As both of these documents used only examples where the sample data appeared to be a good visual fit to a Poisson model, there were no contrasting examples presented to demonstrate how a decision of “poor model”
would be made. Neither document demonstrated calculating and comparing probabilities that combined more than one outcome nor demonstrated how to make an overall judgement as to whether the Poisson model is a good fit to the sample data based on the comparison of individual model probabilities and sample proportions.

After presenting an overall conclusion that the Poisson model was a good fit to the sample distribution, both of the documents (Growing Minds, 2015; Lakeland & Nugent, 2016) also identified individual outcomes within the sample distribution with proportions that did not match the model probabilities. When these outcomes were identified and described, there was no discussion as to why this was done, as the extracts below show:

…. the distribution is similar but $P(X = 5)$ is reduced (Lakeland & Nugent, 2016, p. 58)

…. the probability for $X = 3$ is a little lower for the model (Growing Minds, 2015, section 25.2)

Again the words “reduced” and “a little lower” are subjective, and do not take into account sample size.

For the document written by Lakeland and Nugent (2016), the authors explained that a limitation to using the Poisson model was that the drive-through can only have a maximum of five cars at any point in time. As the variable being modelled cannot take on values above five, how the Poisson model is used would need to be adjusted for this situation, for example, by conditioning all model estimates for probabilities on the event that at most five cars are in the drive-through lane. Instead, the authors wrote that the limit on how many cars can be in the drive-through lane would “…distort the data” (p. 58) rather than discuss the impact of this constraint on the model.

**Using simulations.** All three of the documents published on Census At School (NZ) make use of simulations as part of teaching about probability distribution modelling. Steel (2013) provides several spreadsheets to accompany her written
material, which allow experimental distributions to be simulated from various models using different sample sizes. Her spreadsheets also demonstrate the effect of different bin widths (class intervals) on seeing shape when data is displayed using histograms. While the spreadsheets allow the visualisation of variation, both within the experimental distributions and between different distributions, there does not appear to be a way to compare the simulated data to either the model probability distribution or sample data collected from the real situation.

Martin (2014) also visualises sampling variation through simulation, using animated GIFs to display different simulated distributions from the same model in quick succession. Alongside the animated GIF of simulated data in one colour, is a graph of the real results using a different colour. The use of two different colours could signal that the data being displayed are different in some way and any judgement made on comparing the distribution of the real results with the distribution of the simulated results, while guided by the amount of sampling variation visible, is still subjective.

Addison (2015) developed a spreadsheet tool that incorporates the use of simulation. The tool is set up to simulate new “samples” from the model distribution, allowing users to change the sample size and the parameters for the model. Similar to how simulation is used by Martin (2014) and Steel (2013), the tool can demonstrate the changes in variation for the experimental distribution with changes in sample size, as well as parameter changes. In Addison’s tool, the probability distribution model is shown on the same graph as the simulated distribution (see Figure 11). It is possible to overwrite the simulated data with “real” sample data, but this does not appear to be intention of the tool, as other distributional features in the tool, such as sample space, are fixed.
While the representation used by Addison (2015) has the benefit of allowing the comparison of two distributions in one place (the probability distribution model and the experimental distribution), there is again the potential for confusion with using the word “experimental”. In this context, the visualisation and use of simulation serve to demonstrate that simulated data generated from a probability distribution model will converge to the model (theoretical) probability distribution. This convergence receives a similar focus in the document produced by Hinchliffe and Priest (2013). In their case, an example is provided that involves comparing the probability distributions produced using two different methods to solve a problem: through simulation or by using theory (see Hinchliffe & Priest, 2013, pp. 213 – 215). The theoretical model used for the problem is the binomial probability distribution model. As the simulation also uses a binomial distribution model, the *distribution of experimental estimates of probabilities* is an approximation for the *distribution of model estimates of probabilities*.

However, when using real data collected from the real situation, the *distribution of experimental estimates of probabilities* will converge to the *distribution of true probabilities* as the number of trials is increased. Therefore, there is a potential issue that students will not be able to separate the real world from the modelling world if real data are not used alongside simulated data.
2.4 Conceptual map of potential issues

Following this review of government, commercial and teacher-published resources, and to further illustrate the potential issues with teaching the method *informally testing the fit of the probability distribution model*, a conceptual map of the knowledge required for AS91586 was developed and is shown in Figure 12.

The map shows all the concepts that come into play when students interact with probability distribution modelling, situated around the three different types of probability distributions referred to in AS91586. These concepts need a framework to connect them and the concepts need to be taken into account when determining how to *informally test the fit of the probability distribution model.*
2.5 Development of research questions

The researcher's hypothesis is that there is a problem with teaching students to informally test the fit of a probability distribution model using the method which is described in the curriculum support material (the Senior Secondary Guide). This problem persisted in the assessment items developed by NZQA for the external assessment of AS91586 and in the material developed and published commercially to support teaching and learning of the curriculum achievement objectives, which AS91586 assesses.

The main problem appeared to be that no objective criteria are used to reach a conclusion as to the "goodness" of the probability distribution model. However, as this problem was identified solely based on the researcher's analysis of the relevant New Zealand materials, there was a need to explore and understand the potential issues further.

The main research question for this study is:

How do teachers informally test the fit of a probability distribution model?

Supporting research questions are:

1. How do teachers understand “the distribution of experimental estimates of probabilities”?

2. How is the informal test of the fit of a probability distribution model taught within the wider probability distribution modelling process?

3. How are statistical inference concepts used when teaching probability distribution modelling?

These research questions were used to guide a review of related international statistics education literature, and to design an exploratory study with teachers familiar with teaching AS91586.
Chapter 3

Literature review

3.1 Introduction

A review of the resources available within New Zealand to support the teaching of curriculum achievement objectives associated with AS91586 in Chapter 2 identified a problem with the teaching of the method *informally testing the fit of a probability distribution model*. The main problem appeared to be the lack of objective criteria to test the “goodness” of the probability distribution model. Specifically, the problem was how to make a call about the similarity or dissimilarity of the features of the model probability distribution and those of the sample (experimental) distribution. Also identified were conceptual conflicts with teaching the method alongside simulation-based statistical inference at curriculum level eight (Year 13) of the New Zealand Curriculum. This chapter reviews the limited statistics education research literature related to teaching the method *informally testing the fit of a probability distribution model*.

3.2 Tests for goodness of fit

The method *informally testing the fit of a probability distribution model* was discussed in Chapter 2 and involved the comparison of the features of the sample distribution and the model probability distribution in order to reach a conclusion about the goodness of fit. In attempting to source examples of research involving an *informal test for the fit of a probability distribution model*, however, it was first necessary to more carefully define the characteristics of an informal test.

3.2.1 Characterising an informal test for goodness of fit

Within probability modelling learning contexts, there are examples of research related to the use of informal inferential reasoning (IIR), specifically studies where students discuss model fit by comparing features of distributions (e.g. Fielding-Wells & Makar, 2015; Konold & Kazak, 2008; Lehrer, Jones & Kim, 2014). However, the learning activities in these research studies do not formally use probability distributions such as the normal, Poisson, binomial, uniform or
triangular distributions as models. Therefore, to focus the literature review it was decided by the researcher that the model being informally tested had to be a specific probability distribution that was known to the teachers and/or students in the study.

One reason for teaching informal approaches to inference is that they can make formal procedures more accessible (Dolor & Noll, 2015). Therefore, it seemed sensible to view an informal test for the fit of a probability distribution model as one which would build understandings and lead towards the use of the chi-square goodness-of-fit test. The chi-square goodness-of-fit test is a formal procedure taught within many introductory-level university courses (Agresti & Franklin, 2014; Peck, 2015; Wild & Seber, 2000; MacGillivray, Utts & Heckard, 2014). The chi-square goodness-of-fit test is concerned with whether the proportions of each group of a single categorical variable follow the model probability distribution. Formally testing the goodness of fit requires the use of statistical inference methods.

Consequently, any informal goodness-of-fit test could also be considered informal inference. This is consistent with Pfannkuch’s (2006) usage of the term informal inference to refer to the act of drawing conclusions from looking at, comparing and reasoning from distributions of data. With additional reference to the Informal Inferential Reasoning (IIR) framework developed by Zieffler, Garfield, delMas and Reading (2008) the researcher used the following definition for an informal test for the fit of a probability distribution model to source and review statistics education research literature:

An informal test for the fit of a probability distribution model is one that:

- does not use formal procedures, methods or language e.g., test statistic, null hypothesis, chi-square, P-value;
- draws a conclusion about the goodness of fit of a probability distribution model by looking at, comparing and reasoning from distributions of data;
- builds conceptual understanding of the goodness of fit of a probability distribution model; and
• provides foundations to make the procedures associated with the chi-
  square goodness-of-fit test more accessible.

Before discussion of the limited existing research on informally testing the fit of a probability distribution model, a brief overview of the chi-square goodness-of-fit test is provided.

3.2.2 The chi-square goodness-of-fit test

Pearson's chi-square goodness-of-fit test is a common statistical procedure used to test the fit of statistical model to data that have the form of a single categorical variable with \( k \) groups. The goodness-of-fit test is a formal test for significance, requiring the use of null and alternative hypotheses, a test statistic and a sampling distribution for the test statistic. In the case of goodness-of-fit tests, the null hypothesis is the proposed probability distribution model. In the chi-square test, the test statistic is defined as:

\[
\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}
\]

The test statistic is a measure of the discrepancy between what was observed and what was expected under the null hypothesis. The test statistic is then compared to the chi square distribution (the sampling distribution) with degrees of freedom \( k - 1 - d \), where \( d \) is the number of parameters estimated in the model using the observed data, to determine the P-value. When considering the shape of the chi-square distribution, the peak is related to the degrees of freedom (Dolor & Noll, 2015). At the 5% level of significance, if the P-value is less than 0.05, a conclusion can be made to reject the null hypothesis (the model distribution), otherwise the null hypothesis cannot be rejected. Rejecting the null hypothesis is to conclude that at least one of the groups has a different proportion from the model. The test requires care with how groups are defined (also referred to as binning) and the size of the expected counts for each group, in particular that all expected counts exceed 1 and most (e.g. 80%) expected counts are at least five.
An example of the calculations associated with performing a chi-square goodness-of-fit test is provided in Table 4. For this test, a suspected “dodgy” die is rolled 60 times and the number that lands face up is recorded. For the example shown in Table 4, the chi-square test statistic is 43.8, the degrees of freedom are 5, and the associated $P$-value is $< 0.0001$.

**Table 4: Example of calculations to support chi-square goodness-of-fit test**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed count</strong></td>
<td>14</td>
<td>8</td>
<td>27</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td><strong>Expected count</strong></td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>$(O - E)^2 / E$</td>
<td>1.6</td>
<td>0.4</td>
<td>28.9</td>
<td>4.9</td>
<td>1.6</td>
<td>6.4</td>
<td>$\chi^2 = 43.8$</td>
</tr>
</tbody>
</table>

Based on the test result, it can be concluded that there is very strong evidence against the die being fair, that the proposed probability distribution model (discrete uniform) is not an appropriate model and that this model is not a good fit for the sample data. The size of the discrepancy between what was observed and what was expected is too large to explain it away as sampling variation. The limited research that has been conducted to explore the use of informal methods for goodness of fit tests is now discussed.
3.2.3 Existing research for informally testing the fit of a probability distribution model

Using the researcher's definition for an informal test for the fit of a probability distribution model (see 3.2.1), two studies were found that explored the teaching and learning of the method: Dolor and Noll (2015) and Roback, Chance, Legler and Moore (2006). Both studies shared a common goal for statistical thinking to drive the learning and for students to better understand the formal chi-square goodness-of-fit test by constructing an informal goodness of fit test. Both sets of researchers created and used a set of test samples with students to help them develop initial ideas about what features of sample distributions would support the model distribution being tested and what features would not, including ranking the test samples. After exposure to these initial ideas, the students were guided to develop their own method for measuring the discrepancy between the model distribution and the sample data, again using the test samples. Sampling distributions for the different test statistics were then generated using simulation and used to provide evidence against the fit of the model for each of the test samples.

Dolor and Noll (2015) introduced an additional teaching strategy. Using an example of a jar that contained equal numbers of four different coloured beans, the researchers asked students to give representative intervals for how many beans of each colour they would expect to observe in samples of 100, for example 22 – 28 for each colour. Students were specifically told to consider the expected variability from sample to sample when creating these representative intervals. The researchers included this task in the learning trajectory to generate discussion around unusualness and to encourage thinking around how to judge sample distributions as being similar or different to the model distribution. However, Dolor and Noll observed that students found it challenging to create the representative intervals, as they had not built up enough experience working with repeated samples.

In terms of the overall goal for students to construct an informal goodness of fit test, Dolor and Noll (2015) found that students were able to reason with the
simulated sampling distribution created from their own measure of discrepancy, and were even able to consider the shape of the sampling distribution in relationship to the degrees of freedom and to the nature of the measure. In contrast, despite seeing value in students developing their own test statistics, Roback et al. (2006) observed that students did not independently make use of the simulated sampling distribution to assess whether the test statistic provided evidence against the fit of the model, needing teacher encouragement to use the sampling distribution.

A feature of both studies was that the use of simulation was restricted to generation of the sampling distribution for the measure of discrepancy or test statistic. Although the construction of the sampling distribution through the use of simulation would show variation from sample to sample, the variation is for the test statistic, which is a measure that is not directly visually connected to the sample distribution and the model probability distribution. Hence a limitation using this teaching approach is that the test statistic condenses a myriad of understandings into what could be viewed as an abstract measure. In comparison, the randomisation test that NZ students experience at curriculum level eight (Year 13), has a test statistic - the observed difference between two means – that is still able to be visually connected to the original samples and the sampling distribution (re-randomisation distribution).

Another limitation of teaching an informal test for the fit of a probability distribution model that uses a single numerical measure as part of the test is that the distributional features of the sample and probability distribution model could be devalued; features such as sample space, number of outcomes/groups, variation of individual proportions, influence of sample size and the interaction between these aspects. Also, the use of probability distributions such as Poisson, binomial, and triangular as models is a new concept for Year 13 students and there is lot to unpack and understand about each distribution from a theoretical perspective. Therefore, when considering an informal test for the fit of a probability distribution model that would be suitable for students working at curriculum level eight (Year 13) of the New Zealand curriculum, it was decided to
pursue the possibility of a test that does not use a single numerical test statistic but instead uses a visual comparison of features of the model and sample distributions.

Both Dolor and Noll (2015) and Roback et al. (2006) used a set of test samples to encourage students to consider informally which samples would support the proposed distribution model and which would not. The researcher believed this approach could be leveraged to create an informal test that did not require a numerical test statistic. This led to the consideration of an informal test that is based on comparisons of shape; shape being an overall visual measure of a distribution. The review of research then investigated studies that explored distributional shape features.

3.3 Comparing distributional shape features
According to Pfannkuch et al. “seeing structure and applying structure are important aspects of probability modelling” (2016, p. 11). If seeing structure is interpreted as producing an image of the distributional shape of observed sample data, then applying structure could be interpreted as selecting and using a probability distribution model. As discussed in Chapter 2, particularly in reference to the assessment item for AS91586 shown in Figure 1, students are expected as part of the method informal testing the fit of a probability distribution model to compare the shape of the sample distribution with the shape of the model distribution. For this to form the basis of a reliable test, there would need to be a reliable method to describe the shape of each distribution.

3.3.1 Existing research for sketching distributional shape
A small selection of textbooks written by US statistics educators contained brief explanations on how to draw or conceive of distributional shape using predominantly histograms to display the distribution. Agresti and Franklin (2014) described shape in terms of making the bars of a histogram smaller and smaller as more data is added to smooth the shape. Peck (2015) focused on the idea of shape as approximating a histogram with a smooth curve. MacGillivray et al. (2014) warned of reading too much into peaks or modes in continuous data due to the
varied ways continuous data can be grouped. All of the textbook authors provided explanations about words used to describe shape: symmetric, skewed, unimodal, bimodal etc., and asked questions concerning the use of these words to describe shape of distributions. However, they did not require students to sketch shapes of distributions.

Arnold's (2013) work appears to be the only substantial research into sketching the shape of distribution. As part of a study with Year 10 NZ students, Arnold conducted research involving students sketching shapes of distributions, linking contextual information and shape to identify variables, and developing appropriate language to classify and describe distributions seen in data. All the examples discussed in her research were for large samples where the representations of the distributions were dot plots and where the variables explored were ones set in contexts the students could have personal knowledge about (e.g. hair length, age of students, age of teachers). To encourage students to sketch smooth distributional shapes, Arnold (2013) used an activity where context-free distributions were projected to students on a screen, each for only a brief amount of time. Students were then asked to sketch the distributional shapes on a blank piece of paper after each one was displayed. The purpose of the activity was to focus students on the signal of the distribution rather than the detail of the distribution. Arnold developed criteria for the quality of a distributional shape sketch, with low scores given for shapes that were over-fitted to the distribution (see Figure 13) and the highest score for shapes that were smooth representative curves of the overall distributional shape.

![Figure 13: An example of an over-fitted distributional shape sketch](image-url)
Arnold (2013) quotes Bakker regarding the nature of signal and noise to justify the focus on students sketching a smooth shape rather than the over-fitted shape, “The signal is the continuous shape with which they model the data, and the noise is the variation around the signal” (as referenced by Arnold, Bakker, 2004, p. 234). Arnold believed that the shape sketched should be an inferred shape of the distribution not the shape of the data distribution at hand. The use of contexts for the distributions was to allow for the contextual information to be used to inform the sketch.

There are possible limitations in using Arnold’s work to inform teaching for probability distributions. Her work was developed within the learning context of sample-to-population inference and a driver of the research was to improve the language students used when describing the shape of real contextually-bound distributions. It was also conducted with students working at curriculum level five (Year 10) of the New Zealand Curriculum, where students are not consciously using models as part of their inferential thinking. The sample distributions that were sketched were also large in sample size with clear distributional shape features. Therefore, further consideration of what the sketch of a sample distribution represents when performing an informal test for the fit of a probability distribution model was needed.

3.3.2 Sketching distributional shape from a modelling perspective

It is unclear in Arnold’s (2013) research whether the sketch of distribution shape is inferred about the real population distribution or about the model for the real population distribution or about both. When learning about sample-to-population inference without the explicit use of models, this lack of clarity may not impact understanding, as the only data being used come from the population. But when engaging with probability distribution modelling, students can work with data that are simulated from the model probability distribution alongside real data observed from the process being modelled.

It also seems problematic to consider contextual knowledge when sketching the shape of distribution from a modelling perspective. Consider the “dodgy” die
example discussed in 3.2.2 and the displays shown in Figure 14. Without using contextual knowledge about dice being fair, a distributional shape curve could be sketched as attempted in Figure 14b. By using contextual knowledge about dice being fair and considering the shape of the underlying distribution, a distributional shape curve could be sketched as attempted in Figure 14c. However, for this problem, the contextual knowledge about a die being fair is the basis of the model probability distribution that is tested.

![Figure 14: Two examples of sketching shape for the same sample distribution](image-url)
The potential for confusion could suggest that a distributional shape sketch for sample data (data collected from the real system or process) should be driven by the features of the data at hand. This would keep the shape of the true unknown distribution separate from the shape of the model probability distribution. The use of context to sketch features of data has also been found to be an issue within other types of modelling. In a small study with teachers concerning simple linear regression modelling, Casey and Wasserman (2015) found that some teachers incorrectly used context to decide where to sketch a line of best fit on a scatterplot produced from sample data. For example, some teachers ignored points that did not fit their preconceived idea of what the relationship should be, or forced the line to go through certain points, based on justifications using contextual knowledge of the situation being modelled.

However, the notion of sketching a distributional shape that is informed by the features of the data distribution only and not contextual knowledge is difficult. Unlike the sketch of a line of best fit, there is not a mathematical formula to follow to guide the placement and shape of the sketch. Additionally, in the cases where a small amount of data is available it would be very difficult to see the signal through the noise. For situations where the data are being collected from a complex system or an unfamiliar situation, there may be no ability to predict or imagine the shape of the true underlying distribution. There is also the nature of the type of shape sketched. Arnold (2013) required that a curve be drawn for the highest score in her criteria for the distributional shape sketch. However, there are probability distributions such as the uniform and triangular that are polygons with straight edges. Additionally, it also needs to be considered whether using curves as sketches for discrete random variables is appropriate or helpful.

Using the shape of the data as a criteria for model fit was explored by Lehrer, Jones and Kim (2014). In their study Grade 6 students were asked to evaluate the goodness of fit of a model composed of fixed signal and random noise. Within a repeated measures learning task, students were challenged to assess and revise a model by considering the shape of the model, the data generated from the model and the real world data. However, most students judged the model fit by
comparing statistics associated with the model and the real world data, such as median or interquartile range. Very few students discussed shape as a criterion for testing the fit of the model.

The over-fitted shape sketch (see Figure 13) was not encouraged by Arnold (2013) in her research, and received low scores according to her criteria for sketching the shape of the distribution of a variable. The potential benefits of over-fitted shape sketches within learning about statistical modelling are now discussed.

3.3.3 The concept of overfitting models

In probability modelling, the concept of overfitting is very important. Overfitting involves creating a model that treats everything in the data – signal and noise – as important information and incorporates these features into the model. A student who overfits the distributional shape demonstrates overfitting the model. For example, in Fielding-Wells and Makar’s (2015) work with young children building ideas of probability models (which the authors called inferring to a model), different students’ sample distributions with over-fitted shape sketches were laid on top of each other. Although individually the over-fitted shape sketches did not show the shape of the underlying distribution, collectively they started to paint a picture. This approach could show students why a good model is unlikely to be an over-fitted shape, thus reinforcing understanding of ideas of signal and noise.

Rather than discourage this shape sketching behaviour, it seems that students should be shown why overfitting the shape is unlikely to show the shape of the true underlying distribution. To keep students’ learning within the real world, multiple (or repeated) samples from the true situation being investigated are needed, similar to the test samples used by Roback et al. (2006) and Dolor and Noll (2015). More common in probability modelling activities is the use of physical or computer models so that experimental distributions can be generated quickly and visually. For example, in a study with young children conducted by Konold and Kazak (2008), the shape of simulated samples from a discrete triangular distribution were compared to the model distribution and students were asked to rate the degree of fit of the expected distribution either as bad, OK, or great. This
activity could help students learn about noise, and how the shape of the small samples may not look like the model.

Important statistical ideas concerning models, overfitting and noise can be found in the research conducted by Zieffler (2016). Zieffler used model eliciting activities (MEAs) with students and informal methods to develop a model to predict if an email was spam or not based on its subject. When discussing important statistical themes that emerged from the learning activity, Zieffler identified that students learned about how different samples of data would give different models and how using the “best” model from a single set of data leads to overfitting the model. In this learning context, the “best” model was one developed that correctly classified every email in the sample data as spam or not using many specific rules (the model), a process that makes the model over-fitted to the sample data. However, the “best” model was based on noise as well as signal, and therefore the model was unlikely to be generalisable.

The process of sketching an over-fitted shape produces a discrete random variable. If this discrete random variable is used as a model, just as in the spam email MEA used by Zieffler (2016), then the issue is not in the process of creating the model but in the lack of generalisability of the model. As students should be learning that modelling, and inference, require going beyond the data, the ideas of overfitting support this greater understanding. The problem is not resolved on whether the shape sketched of a sample distribution should be driven by the features of the data at hand, or by contextual knowledge of the process creating the data, or even by the model being considered for the process creating the data. This problem is not unique to probability distribution modelling. Casey and Wasserman (2015) explored teachers’ conceptions for the line of best fit. In their small study of 19 teachers, there was a near even split between those teachers who saw the line sketched as a model for the whole population or a signal which could be used to predict values for those not in the set \( n = 9 \) and those teachers who saw the line of best fit as typical or representative of the sample data only \( n = 10 \).
A final comment on the benefit of teaching students about overfitting and the relationship of this to sketching over-fitted shapes is the potential of the visualisation of these over-fitted shapes. Within a software driven modelling environment, if the over-fitted distributional shapes were able to be tracked from simulated sample to simulated sample from the model, then this would demonstrate that noise is variation around the signal. This strategy holds promise for use within the development of a new informal test for the fit of a probability distribution model.

3.4 **Towards a new informal test for the fit of a probability distribution model**

There was limited research literature available regarding informal tests for the fit of a probability distribution model that incorporated aspects of informal inferential reasoning (IIR). However, the literature that was reviewed proved informative and useful to the development of a new informal test for the fit of a probability distribution model. Dolor and Noll (2015) and Roback et al. (2006) demonstrated how test samples could be used with students to build ideas testing the fit of a probability distribution model. By asking students to look at features of sample distributions, and to compare these to the proposed model distribution, students could develop ideas of how to measure discrepancy between what was expected and what was observed for the chi-square goodness of fit test. However, as Dolor and Noll (2015) noted, students had not built up enough experience working with repeated samples to be able visualise representative intervals for the expected variation of each proportion in the model. A way to target the development of this understanding was shown by Fielding-Wells and Makar (2015), where a single plot, with over-fitted shapes from student samples overlaid on top of each other, could demonstrate sample to sample variation in distributional shape.

The visualisation of over-fitted shapes from repeated samples appeared to the researcher to be a strategy that could be used as part of teaching students how to sketch the shape of a distribution, although questions concerning what a distributional shape sketch should represent and whether contextual knowledge should be used were unresolved. Moving to the model world, Konold and Kazak
(2008) and Lehrer et al. (2014) used simulation to explore the relationship between the shape of the model distribution and the shape of data simulated from the model, and encouraged students to use shape as one of their criterion for assessing model fit. Dolor and Noll (2015), Lehrer et al. (2014) and Roback et al. (2006) also used simulation to develop sampling distributions for student-created test statistics (measures of discrepancy), although the researcher decided that the use of single numerical test statistic could require thinking beyond what is required or reasonable for students working at curriculum level eight (Year 13) of the New Zealand Curriculum. The success of visual-based and simulation-based methods in the research reviewed led the researcher to review new research related to simulation-based visual inference.

### 3.4.1 Simulation-based visual inference

Students working at curriculum level eight (Year 13) of the New Zealand Curriculum are expected to be able to do a simulation-based randomisation test to “make a call” about the difference between two means (or medians) using Chris Wild’s VIT (Visual Inference Tool)\(^8\). A screenshot of the VIT randomisation test module is shown in Figure 15.

New research for simulation-based inference (SBI) is emerging (e.g. Chance, Wong and Tintle, 2016; Noll, Gebresenbet & Glover, 2016). A common feature of SBI is the creation and use of a numerical test statistic and formal procedures such as the definition and use of hypotheses. The use of simulation is also primarily geared towards the creation and use of a sampling distribution for the test statistic. Alternatively, Hofmann, Follett, Majumder and Cook (2012) developed a visual inference method (also called graphical inference) where data plots are used as test statistics.

---

\(^8\) See [https://www.stat.auckland.ac.nz/~wild/VIT/](https://www.stat.auckland.ac.nz/~wild/VIT/)
A lineup of plots is generated, consisting of the real sample data plot randomly placed somewhere between plots generated from the null hypothesis or a known model. If someone viewing the lineup of plots can identify the real sample data plot, then that gives statistical evidence to support a conclusion that the real sample data does not fit with the null data (the model tested). An example created by the researcher to demonstrate this approach to testing the fit of a probability distribution model is shown in Figure 16. The real sample data are the 60 rolls of the “dodgy” die from the example discussed in 3.2.2 (see Figure 16). The visual inference method used by Hofmann et al. (2012) could allow students to draw a conclusion about the goodness of fit of a probability distribution model by looking at, comparing and reasoning from distributions of data. However, it lacks the visual animation built into Wild's VIT modules.
Any informal test for the fit of a probability distribution model needs to take into consideration a framework for its use. In regards to the definition created by the researcher in 3.2.1 as to the required characteristics of the informal test, one of the necessary characteristics was that the test would build conceptual understanding of the goodness of fit of a probability distribution model. To build conceptual understanding requires use of an informal statistical inference method such as the one demonstrated by Hoffman et al. (2012) within the teaching of probability distribution modelling. Therefore, the researcher developed a statistical modelling framework that could assist with teaching probability distribution modelling alongside statistical inference.
3.4.2 Proposed statistical modelling framework

As discussed in Chapter 2, there may be a potential issue with how teachers conceive the relationship between distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities. In considering how to integrate statistical inference with probability distribution modelling in the development of a statistical modelling framework, it was hoped that this issue could be resolved.

In reviewing research literature, the earliest work the researcher could source that discussed the relationship between true, model and experimental or empirical probability was a paper by Konold et al. (2011). The paper is based on a case study of one student and discusses some important ideas about how to conceive the three types of probability distributions. In the case of experimental probability, textbooks often define this as a single probability, based on long run relative frequency calculated from a large number of trials, and further define that the experimental probability is an estimate for the model probability. Konold et al. argue it is necessary to conceptualise experimental probabilities for the same event as estimates of the true probability. They further explain that this relationship is obscured when situations are used where the model probability and the true probability are the same (or very close to being the same).

Pfannkuch and Ziedins (2014) developed a diagram to demonstrate these three interconnected ways to think about probability, shown in Figure 17. Although the diagram does clearly show that experimental probabilities (referred to as empirical probabilities in Figure 17) are used to estimate the true probability, there is also an arrow connecting model probability and experimental probability. In considering how to integrate statistical inference into this thinking, it seems sensible and obvious to consider sample data as experimental data, as probability estimates calculated from sample data can be used to estimate the true probability, just as sample means can be used to estimate the true population mean.
Although research literature concerning distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities is scarce, these three ways of thinking about probability are not new to probabilists. Within a 62-year-old report on the Use of Poisson distribution in highway traffic (Gerlough & Schuhl, 1955) under the heading Testing Goodness of Fit (χ² Test) the following was written (pp. 8 – 9):

In each of the foregoing examples it has been postulated that a Poisson distribution having a parameter \( m \) whose value has been computed from the observed data describes the population that has been sampled. The observed distribution constitutes this sample. By inspection there is apparent agreement between the observed distribution (sample) and the theoretical distribution. The inference is then made that the postulated theoretical (Poisson) distribution is in fact the true population distribution. This inference is based, however, solely on inspection; a more rigorous basis for reaching such a conclusion is desired. One of several statistical tests of significance may be used for this purpose; the chi square (\( \chi^2 \)) test is appropriate to the present application. This test ... provides for one of two decisions:

1. It is not very likely that the true distribution (of which the observed data constitute a sample) is in fact identical with the postulated distribution.
2. The true distribution (of which the observed data constitute a sample) could be identical with the postulated distribution.

This example clearly demonstrates the three ways of thinking about probability from a modelling perspective, and shows how statistical inference can be incorporated into this perspective. However, when considering how to integrate the use of simulation based inference into this perspective rather than the chi square goodness-of-fit test, the researcher believes this perspective has limitations. In modelling activities that involve both real data and simulated data, such as simulation-based inference, students and teachers can confuse real distributions with simulated distributions (Gould, Davis, Patel, & Esfandiari, 2010; Pfannkuch, Wild, & Regan, 2014). This suggests that simulated data need to be treated differently from observed sample data, especially when teachers also need to consider how simulation introduces another source of variation for the data (Engel, 2010; Watkins, Bargagliotti, & Franklin, 2014).

There is also the potential for confusion of statistical ideas concerning inference when using data as part of teaching probability. In his work with using probability experiments to build ideas of inference, Nilsson (2014) presented the components of this inference direction as experiment -> empirical probability -> theoretical probability. In this definition of inference, students used the data collected from a probability experiment to form an estimate (an empirical probability) of a model or theoretical probability, or as Nilsson explained “...to make inference from data to the features of the random generator producing the data” (p. 512). Fielding-Wells and Makar (2015) also refered to inferring to a model in their probability modelling research, and used data generated through experimentation to help students build ideas of probability models. In both the aforementioned studies, the experimentation involved drawing objects from bags at random where the probability models were unknown to the students. Therefore, the researcher argues that the students are using the data generated from these models to make an inference about the true and unknown probability, not directly about the model probability. That is, students’ reasoning process needs to follow the route from sample data to true probability to model probability. The reason why both of these
examples of probability modelling can also validly infer to a model probability is because the situation explored is one where the model is the same as the truth. However, in other situations, the model may not be the truth, only an approximation of the truth. In these other situations, the data collected from the real situation can only be used to infer to the true probability and not the model probability, as the data are not generated from the model.

Therefore, the researcher proposes the statistical modelling framework shown in Figure 18.

![Figure 18: Key components of the proposed statistical modelling framework](image)

The framework consists of two parallel worlds: the real world and the model world. In both worlds, data are the output of a process that involves randomness, and the resulting data can be used to learn about this process. The framework allows for the separation but connection of data that are observed in the real world and data that are generated from a model, and the separation but
connection of the true unknown random process that is being modelled and the model itself. The purpose of the framework is to clearly show the different components of statistical modelling, to differentiate between the real world and the model world, and to theorise how one might move between the two worlds when engaging in modelling. The framework consists of four components: The real/true (unknown), real data, the model (known), and model data. Simulated data are a type of model data. The proposed framework therefore allows for separate and connected thinking about probability in the real world and the model world. Hence this framework will be used to help answer the research questions for this study.
Chapter 4
Methodology

4.1 Introduction

This research is a small-scale study, allowing for both exploration and confirmation of conjectures. Exploratory studies are associated with qualitative research, where the priority for the research is to collect rich information not to obtain a representative sample nor to generalise findings to a wider population (Creswell, 2015; Creswell, 2002; Patton, 2015). Qualitative data consists of words, actions, or other observable data that are not recorded as numbers. However, it is possible to transform qualitative data to quantitative data through coding systems that allocate numerical scores on a defined scale. Although quantitative data analysis is associated with confirmatory studies, it is possible to collect and analyse quantitative data within a study framed by qualitative methods. One way to do this is through using a mixed methods design.

4.2 Mixed methods research

Creswell (2015) defines mixed methods as “an approach to the research in the social, behavioural, and health sciences in which the investigator gathers both quantitative (close-ended) and qualitative (open-ended) data, integrates the two, and then draws interpretations based on the combined strengths of both sets of data to understand research problems” (p. 2). This approach requires the researcher to make decisions about the research procedures from philosophical (worldview) and theoretical perspectives (Creswell & Plano Clark, 2011; Punch & Oancea, 2014). A worldview or paradigm associated with mixed methods research is pragmatism. From a pragmatist worldview, researchers are free to do “what works” to answer their research questions, and can draw on multiple ideas, beliefs, methods and concepts of knowledge (Creswell & Plano Clark, 2011). A pragmatic approach is practical in orientation, and focuses on finding a solution to a problem that fits a defined context without holding on to preconceived ideas or principles, and asks questions like What is the consequence of doing this? (Biesta, 2012; Hammond & Wellington, 2013; Morgan, 2014).
A general rationale for using mixed methods research is that the strategy allows the researcher to learn more about the phenomena of interest, as the strengths of both quantitative and qualitative research methods and analysis are combined and their respective weaknesses compensated for in the process (Biesta, 2012; Creswell, 2015; Punch & Oancea, 2014). The justification for the use of mixed methods by researchers is based on an assumption that the method will best answer their research questions (Creswell & Plano Clark, 2011), and that the use of only quantitative or qualitative research methods alone would be insufficient (Creswell, 2015). For example, the use of qualitative interviewing followed by experimental intervention within the same study would allow the researcher to create a theory and then test that theory. The mixed methods research design allows the researcher to generate interpretative understanding using qualitative data that captures why people think and act a certain way, and uses the quantitative data to make this understanding more robust (Biesta, 2012).

Researchers using a mixed methods design need to be aware of the different aspects of research that can be mixed or combined. Biesta (2012, p. 148) provides seven questions to assist in researchers’ reflections, which include the questions: “Is it possible to combine text and numbers?”, “Is it possible to combine different methods of data collection and/or data analyses?” and “Is it possible to combine experimental/interventionist and naturalist/non-interventionist designs?”. As stated earlier, the key driver of the mixed methods approach are the research questions. However, whether it is realistic to adopt a perspective that research questions should determine research methods has been challenged by Morgan (2014). Morgan suggests that it is better to think of research as a cyclical process, where the researcher’s personal interests are a major factor for decision making, and where choices about research questions can be influenced by choices about research methods.

Common instruments to collect data in mixed methods research are questionnaires, interviews and classroom observations (Zohrabi, 2013). Within education research, the use of an interview to collect data is a prominent tool (Punch & Oancea, 2014).
4.3 Use of interviews

An interview can be used to explore participants’ views, experiences, meanings, thoughts, values, feelings and perceptions about a phenomenon (Hammond & Wellington, 2013; Punch & Oancea, 2014). Within an education research context, an interview can support data collected through assessment tasks and allow the researcher to learn more about each participant’s reasoning. However, what participants say in an interview may not be what they actually do and how they actually think, and so researchers who use interviews have to assume that participants are aware of themselves for their answers to represent reality (Powney & Watts, 1987).

Different types of interviews include structured, focused or semi-structured, and unstructured. Research methodology literature tends to categorise an interview as a qualitative data collection method, or uses a more specific description of a qualitative interview. However, a structured interview can also be viewed as a questionnaire or survey, for example a face-to-face questionnaire (Hammond & Wellington, 2013). An advantage of using a structured interview is that all participants received the same questions in the same way and this consistency in participant experience can aid with reducing bias in the results, for example interviewer bias. However, due to the constraints on what can be asked, a major limitation of a structured interview is that there is no conversation between the interviewer and the interviewee, which means that answers that may contain interesting ideas that could be explored by the interviewer are not responded to with follow-up questions.

Interviews can be conducted face-to-face, over the phone, or through internet-based communication tools such as Skype, Facetime and email. While research methodology literature tends to categorise a survey as a quantitative data collection method, a survey can be considered as a self-interview as each participant is in effect interviewing themselves (Creswell, 2002). Internet-based interviews can allow the researcher access to participants that are located in other areas of their country or the world and in different time zones. Additionally, for some research contexts, the environment within which the participant completes
the interview may be less intrusive and allow for more time and opportunity to reflect on answers (Hammond & Wellington, 2013). Possible issues with internet-based interviews are that the researcher will not be aware of the environment within which the participant has completed the interview and whether the responses are from the participant or someone else.

Interviews are developed by creating a set of questions (an interview schedule) that reflect the nature of the research, that will inform the research questions, and that will also be meaningful for the interviewees. According to Patton (2015), there are six types of question that can be asked: experience and behaviour questions, opinions and values questions, feeling questions, knowledge questions, sensory questions and background or demographic questions. All of these questions can be asked of participants based on the past, the present and the future. Interviews tend to comprise mostly open-ended questions, however, other types of question formats can be used within an interview. The nature of open responses is that they will need to be analysed to create meaning. This relies on the researcher's integrity and skill in qualitative data analysis, in particular to include all relevant data whether it supports the researcher’s point of view or position or not (Powney & Watts, 1987).

For task-based interviews, interviewees are asked to complete a task and often asked to verbalise their thoughts while or after completing the task. Questions asked tend to be designed to elicit knowledge, but can also be used to ask about experiences, opinions and feelings. For example, in a study with teachers from the United States of America, Casey and Wasserman (2015) used a task-based interview to measure teachers’ understanding of informal lines of best fit. One task required teachers to fit lines of best fit for different sets of bivariate data, and to describe what they were doing and why throughout the model fitting process. Watson (2001), in a study with teachers from Australia, included tasks designed to elicit teacher views and knowledge. Tasks included asking teachers to brainstorm factors considered important for the teaching of a topic, and describing an activity to teach a specific word such as “sample”.
The data collected from interviews generally are qualitative, and therefore require the use of qualitative analysis methods to explore and find meaning in the words and actions used during the interview.

4.4 Qualitative data analysis

According to Braun and Clarke (2006), thematic analysis can be viewed as a foundational method for qualitative analysis and should be the first method qualitative researchers learn. The goal of thematic analysis is to identify repeated patterns of meaning across a data set, which are called themes. Braun and Clarke present guidelines for thematic analysis, organised by the following six phases: (1) familiarise yourself with data, (2) generate initial codes, (3) search for themes, (4) review themes, (5) define and name themes, and (6) produce the report. Braun and Clarke propose that thematic analysis is an accessible and theoretically flexible method for analysing qualitative data.

Qualitative data can also be analysed using structural analysis. In this approach data are compared to a pre-existing structure or framework for characterising processes, actions or behaviours, and a system is developed that allows for data to be encoded in a diagram which structures the participants’ reasoning according to the framework (Grigoraş, Garcia, & Halverscheid, 2011). For example, Grigoraş et al. used the words spoken by students during a modelling activity to align their actions to a mathematical modelling framework. The framework allowed the researchers to visualise transitions between the real world and the mathematical world along a time scale. As with thematic analysis, the coding of actions, either observed or described, is determined by the researcher, which requires a clear understanding of the structure or framework being used. This also requires the researcher to decide how to code the actions, which can be a subjective decision if there are not clear criteria.

The review of this related research methodology literature informed the research design for this study, which is now presented.
4.5 Research design

4.5.1 Mixed methods design

A mixed methods design was used for this study as the research questions could be better answered using a combination of quantitative and qualitative methods. Much richer answers to the research questions were possible than if only one type of method and analysis was used.

From within a pragmatic worldview, the researcher used the research questions to select methods for data collection and analysis based on their suitability and effectiveness to answer the research questions. An embedded design was used for this mixed methods research. An embedded design is a type of design in which a quantitative method is embedded within a qualitative data collection method. Specifically, for this study a randomised experiment was used within a task-based interview. This design allowed for both exploration and confirmation of the theory related to the research questions.

The diagram of procedures used is provided in Figure 19 and each procedure is now explained in more detail.

Figure 19: Diagram of procedures for this study
4.5.2 Selection of participants

The participants in the study were 17 mathematics and statistics teachers from across a range of New Zealand high schools. Purposeful sampling was used to obtain participants for the study, as teachers needed to have had experience teaching certain achievement objectives from the New Zealand Curriculum (Ministry of Education, 2007). The researcher placed an advertisement on the NZ statistics teachers Facebook page, and asked teachers to share this advertisement with other teachers who might be interested. Teachers who contacted the researcher were checked for teaching experience with AS91586, then provided with the participant information sheet and consent form (see Appendix B) through email.

In total, 28 teachers made contact with the researcher, 20 returned the consent form, and 17 teachers completed at least one task beyond the first task. These 17 New Zealand statistics teachers had all taught or were teaching AS91586. Four of these teachers had completed an undergraduate degree with a Statistics major or equivalent. On average, teachers had taught for 13 years (median = 13, mean = 16.2), with teaching experience between five years and 40 years. Eleven of the 17 teachers were female and all but one of the 17 teachers had taught AS91586 for at least two years.

4.5.3 Data collection methods

Data were collected from participants during a four-week period over May to June 2016 and through a task-based interview conducted in an online environment. An online environment was chosen for conducting the research to allow for participation of teachers in the project from a wide range of locations within New Zealand. Additionally, some of the tasks in the interview required the use of the researcher’s new computer-based tools for exploring probability distribution modelling that were not available publicly. The online environment was designed and coded by the researcher specifically for this project. A screen shot of the welcome page of this environment in shown in Figure 20.
4.5.4 Task-based interview design

The task-based interview consisted of a series of six new tasks designed by the researcher. In line with the use of mixed methods, the nature of each task was guided by the research questions. Table 5 gives a summary of each task, linked to the research questions and research that influenced the design of the tasks.

<table>
<thead>
<tr>
<th>Task summary</th>
<th>Research questions</th>
<th>Research base</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task B</td>
<td>Preparing to teach AS91586</td>
<td>SQ1*, SQ2</td>
</tr>
<tr>
<td>Task C</td>
<td>Sketching shapes of distributions</td>
<td>SQ3</td>
</tr>
<tr>
<td>Task D</td>
<td>Assessing student understanding</td>
<td>RQ**, SQ3</td>
</tr>
<tr>
<td>Task E</td>
<td>Capturing modelling processes</td>
<td>RQ, SQ2, SQ3</td>
</tr>
<tr>
<td>Task F</td>
<td>Using a new simulation-based tool for probability distribution modelling</td>
<td>SQ2, SQ3</td>
</tr>
</tbody>
</table>

* SQ1 – supporting question 1, ** RQ – research question
The order of the tasks was determined to minimise the effect of providing new information or understanding to teachers through completing the tasks and to try to elicit teacher current understanding prior to completing the task. This was not completely avoided as each task was provided on the main screen with short summary as to its contents. At the end of each task, a final question was used to allow teachers to make any further comments about the task, which could include clarifying any responses or further thoughts/comments related to the task questions. All six tasks are included in Appendix A and key aspects of Tasks B, C, D, E and F are now discussed.

**Task B – Preparing to teach AS91586**
Task B was influenced by tasks used by Watson (2001) in her profiling of teachers’ knowledge. Teachers were asked to describe three key ideas, concepts or understandings that they thought were important for the teaching of lessons related to probability distribution modelling (as assessed by AS91586). They were also asked to identify two aspects that students struggled with when learning about probability distribution modelling. Teachers were also asked to describe an activity that they had used to build understanding of *the distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities*, and to explain why this activity built understanding. Task B was designed to measure teacher understanding of the three types of probability distributions, beyond definitions of each term.

**Task C – Sketching shapes of distributions**
Task C was based on tasks developed by Arnold (2013) regarding the sketching of the shape of a distribution. Teachers were shown six different distributions, each presented with no context. The data for these distributions were sourced from the 2015 Auckland Marathon, and the variable shown in the times to complete the marathon. The different distributions were created by combining or separating the times to complete the marathon by event (full marathon, half marathon), gender and age group. These distributions vary by at least two factors: sample size and potential modality. This was intentional, so that comparisons could be made
between the sketches made for each of these distributions. Table 6 shows each of the six distributions used, cross-classified by these two factors. The number of the distribution indicates the order the distributions were shown to teachers.

Table 6: The six distributions used in Task C, cross-classified by sample size and potential modality

<table>
<thead>
<tr>
<th></th>
<th>Potentially unimodal</th>
<th>Potentially bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small sample size</td>
<td>Distribution 4</td>
<td>Distribution 2</td>
</tr>
<tr>
<td>(n &lt; 30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium sample size</td>
<td>Distribution 6</td>
<td>Distribution 3</td>
</tr>
<tr>
<td>(30 ≤ n ≤ 100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large sample size</td>
<td>Distribution 5</td>
<td>Distribution 1</td>
</tr>
<tr>
<td>(n &gt; 100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using a new online tool developed by the researcher, each teacher was asked to sketch the shape of the distribution using a mouse or their finger if using a touchscreen (e.g. an iPad). The shape sketched for each distribution was captured digitally for each teacher. Each teacher could redraw the sketch if they wanted to, but could not go back to an earlier sketch once they had moved on to the next distribution. This was to prevent teachers redrawing their sketches if subsequent distributions prompted them to review their strategy for sketching.

At the end of Task C, teachers were asked to make any comments about the task. This was included to allow an opportunity for teachers to communicate any difficulties they found with sketching the shapes using the online tool and also to allow teachers to explain why they did not sketch a shape for any particular distribution. Task C was designed to explore what features in a distribution did or
did not influence each teacher's interpretation of its shape, by attempting to capture digitally what shape each teacher saw in the data distribution.

Task D – Assessing student understanding
For Task D, teachers were given a problem concerning whether a normal distribution model was a good model for the lengths of jelly beans and shown two different hypothetical student responses to the problem. The problem involved a small set of sample data ($n = 50$) displayed as dot plot with a box plot underneath (see Figure 21). The data used for this task was real data collected by the researcher. Note a similar assessment item occurred in the NZQA, 2015, AS91586 exam.

Both student responses demonstrated at least one misunderstanding, and teachers were asked to identify the misunderstanding(s) and explain what further explanations and activities they would use to help each student develop the correct understanding.

![Figure 21: Sample distribution used for Task D](image)

The response provided for Student A was as follows:

Student A sketches the shape of the sample distribution of jelly beans lengths (negatively skewed) and compares this to the shape of the model normal distribution (symmetric). As the shapes are not similar, the student concludes that the normal distribution is not a good model for the lengths of jelly beans produced by the confectionery company.

The response provided for Student B was as follows:
Student B uses the sample data to calculate the proportion of jelly beans in the sample that are less than 19 mm in length. The student then uses the model defined above to calculate the proportion of jelly beans that could be expected to be less than 19 mm in length. As the proportions are similar, the student concludes the normal distribution is a good model for the lengths of jelly beans produced by the confectionery company.

Task D was designed to measure teachers understanding of the issues associated with the two approaches currently used to assess the fit of a probability distribution model – comparing shape of the experimental and theoretical probability distribution and comparing theoretical and experimental probabilities – and to elicit examples of activities or explanations they would use to address the misunderstanding(s) demonstrated by each student.

**Task E – Capturing each teacher’s modelling process**

Task E was based on the NZQA assessment item discussed in Chapter 2 and shown in Figure 22. The overall objective for this task was to investigate whether the number of emails received per hour to an email account could be modelled by a Poisson distribution. The first part of the task asked teachers to describe the steps they would take to complete the investigation. The second part of the task directed teachers through an investigation that required sketching the shape of the sample distribution and informally testing the fit of a probability distribution model. Teachers were also asked during the second part of this task to explain their reasoning. This aspect of the task design was influenced by Casey and Wasserman (2015) and their work with exploring teacher understanding of lines of best fit.

Task E was a randomised experiment and two versions of this task were used in the study: Version 1, which provided teachers with a small set of data \((n = 48\), see Figure 22a and 22c); and Version 2, which provided teachers with a large set of data \((n = 480\), see Figures 22b and 22d).
The number of emails received each hour to an email account was recorded over a 48-hour period.

This data recorded is shown below:

| 4, 3, 1, 0, 0, 0, 2, 2, 0, 1, 2, 2, 2, 1, 1, 7.4, 0, 3, 4, 2, 4, 2, 1, 0, 0, 2, 0, 1, 0, 0, 0, 1, 3, 3, 3, 2, 3, 0, 2, 2, 1, 3, 4, 2, 1, 2 |

You have been asked to investigate whether the number of emails received per hour to this email account can be modelled by a Poisson distribution.

(b)

The number of emails received each hour to an email account was recorded over a 480-hour period.

This data recorded is shown below:

| 4, 0, 2, 1, 4, 0, 3, 1, 3, 0, 2, 4, 2, 3, 2, 0, 1, 0, 1, 0, 3, 4, 0, 1, 0, 2, 1, 1, 0, 2, 4, 0, 4, 0, 0, 2, 4, 2, 0, 3, 2, 1, 4, 0, 0, 1, 1, 0, 3, 2, 3, 2, 2, 1, 1, 0, 0, 3, 2, 2, 0, 0, 0, 2, 2, 1, 7, 1, 2, 1, 3, 3, 1, 2, 1, 2, 3, 0, 0, 2, 1, 0, 2, 3, 4, 3, 3, 2, 4, 3, 1, 0, 3, 3, 1, 2, 1, 1, 0, 7, 2, 3, 3, 0, 0, 0, 2, 4, 2, 2, 1, 3, 4, 0, 3, 3, 0, 3, 3, 4, 2, 1, 3, 1, 2, 0, 0, 1, 2, 2, 1, 2, 0, 0, 3, 2, 2, 2, 2, 0, 2, 1, 1, 3, 1, 4, 1, 0, 1, 2, 4, 2, 2, 0, 4, 2, 1, 0, 1, 2, 1, 2, 0, 1, 0, 4, 0, 0, 3, 3, 0, 0, 2, 1, 2, 0, 1, 4, 0, 0, 3, 1, 1, 1, 0, 0, 1, 1, 0, 2, 4, 2, 1, 0, 3, 3, 1, 3, 4, 2, 2, 0, 2, 1, 0, 2, 4, 2, 1, 0, 2, 0, 3, 0, 2, 7, 4, 2, 2, 0, 0, 2, 2, 0, 7, 2, 0, 2, 2, 0, 1, 0, 4, 2, 0, 1, 1, 2, 3, 4, 4, 1, 0, 2, 3, 4, 2, 2, 2, 1, 1, 4, 0, 1, 2, 0, 3, 3, 1, 4, 2, 2, 0, 0, 1, 2, 2, 2, 2, 1, 2, 3, 3, 0, 0, 3, 1, 4, 0, 4, 2, 0, 0, 0, 2, 0, 2, 2, 1, 0, 1, 2, 2, 0, 7, 0, 1, 2, 4, 1, 0, 2, 1, 0, 2, 1, 2, 0, 4, 2, 3, 2, 0, 4, 3, 4, 4, 3, 1, 0, 1, 3, 3, 0, 2, 1, 0, 0, 4, 2, 1, 0, 3, 4, 2, 2, 0, 1, 1, 1, 4, 3, 0, 0, 0, 4, 0, 7, 0, 0, 4, 1, 3, 0, 4, 1, 2, 2, 3, 0, 0, 0, 0, 2, 7, 3, 0, 4, 3, 3, 3, 3, 1, 1, 0, 3, 1, 2, 4, 2, 2, 0, 1, 4, 0, 1, 4, 0, 3, 7, 0, 0, 4, 2, 0, 2, 1, 1, 1, 7, 0, 3, 4, 3, 2, 2, 0, 3, 0, 2, 3, 1, 2, 2, 1, 2, 0, 4, 2, 3, 3, 1, 4, 4, 0, 2, 0, 3, 4, 3, 0, 3, 4, 0, 0, 0, 0, 0, 0, 1, 0, 2, 2, 2, 0, 1, 2, 4, 3, 3, 4, 1, 1, 0, 1, 2, 2 |

You have been asked to investigate whether the number of emails received per hour to this email account can be modelled by a Poisson distribution.

(c)

Version 1, \( n = 48 \)

No evidence against Poisson distribution

with \( \lambda = 1.77 \)

\( (\chi^2 = 5.82, \ p = 0.213) \)

(d)

Version 2, \( n = 480 \)

Very strong evidence against Poisson distribution

with \( \lambda = 1.77 \)

\( (\chi^2 = 156.53, \ p < 0.001) \)

Figure 22: Key design features of Task E
All questions asked in Task E were identical for both versions and the sample distributions used (shown in Figures 22c and 22d) were identical in terms of outcomes (sample space) and proportions (weightings). Teachers were randomly allocated to one of the two versions of Task E when they reached this task. During the task, teachers were also shown the sample data with the theoretical model (Poisson distribution) overlaid (see Figures 22e and 22f) and asked to discuss the appropriateness of the Poisson model in terms of the visual fit of the model to the sample data. For Version 1 the Chi-square test gave no evidence against the Poisson distribution being a good fit for the sample distribution, while for Version 2 the Chi-square test gave very strong evidence against the Poisson distribution being a good fit for the sample distribution.

Task E was designed to confirm the conjecture that teachers do not take into account sample size when assessing the fit of a probability distribution model to sample data and to explore what criteria and reasoning teachers use to assess the fit of a probability distribution model.

**Task F – Using a simulation-based tool for probability distribution modelling**

Task F used a new simulation-based modelling tool that was designed by the researcher for this project. The tool was developed in an attempt to address some of the potential issues identified when reviewing documents and literature (see Chapters 2 and 3), namely: sample data collected from the real situation being confused with simulated data generated from the model; and sample size not being taken into account when informally testing probability distribution models.

One of the innovative features of the tool is that the interface is based on the researcher’s proposed framework for statistical modelling. Therefore, when learners interact with the tool the left hand side of the screen always displays information and data related to the real situation being modelled, and the right hand side of the screen always displays information and data related to the model being used. One of the other innovative features of the tool is the tracking of the over-fitted shape for the simulated data from the model. This allows the learner to visualise the expected variation in shape using animation (using the simulated
samples of the same size as the real data) and to transfer that visualisation to the sample distribution in order to informally test the fit of a probability distribution model. Figure 23 shows screenshots of the three key stages of using the tool to informally test the fit of a probability distribution model.

For Task F, teachers were first guided through how to use the simulation-based tool using two contrasting examples. As teachers moved through this exploration, text appeared in a highlighted box first at the top of the screen, followed after a short delay with demonstrations of the tool in action (see Figure 23). The instructions provided in this guided exploration often combined technical instructions with necessary modelling knowledge. Teachers could not interact with the tool until the end of the guided exploration.

After the guided exploration of the new simulation-based modelling tool, teachers were given the sample data they were presented with in Task E and allowed the opportunity to use the tool to assess the fit of the proposed Poisson model. Teachers were then asked questions about what benefits (if any) they believed using the tool could have for building students’ understanding of informally fitting a probability distribution model, including what key concepts, ideas or understandings the tool could highlight. These questions were used to further elicit understanding from teachers regarding how simulations could be used to build understanding of probability distribution modelling, and to allow for a comparison of what was intended to be the benefits of the tool and what was perceived by teachers as the benefits. Teachers were also asked what understandings about informally fitting a probability distribution model, if any, the tool had helped clarify for them.
Figure 23: Screenshots demonstrating the use of the tool to informally test the fit of a probability distribution model.
4.5.5 Analysis methods

As the tasks used in the task-based interview comprised open ended questions, most of the data collected was of a qualitative nature. Data from each task were analysed separately and no attempt was made to analyse responses from teachers across more than one task. Within individual tasks, responses from teachers were reviewed and initially coded in an attempt to attach meaning to the data (Punch & Oancea, 2014). Thematic and structural analysis methods were used flexibly by the researcher to seek answers to the research questions. Details of the analysis methods used for each task are included in Chapter 5.

4.6 Validity, reliability, ethics and researcher bias

For the results of this research to be taken seriously, the research methods used should be valid and reliable. Within a mixed methods research strategy, Zohrabi (2013) defines validity as whether the research “... is believable and true and whether it is evaluating what it is supposed or purports to evaluate” (p. 258). Zohrabi argues that in terms of reliability “... the purpose is not to attain the same
results rather to agree that based on the data collection processes the findings and results are consistent and dependable” (p. 259). The validity and reliability of this research are now discussed, followed by discussion of the ethics of this study and the potential for researcher bias.

**Validity and reliability**

All tasks used in this research were designed by the researcher and so were not previously externally validated. However, several approaches were used to ensure the validity of the tasks used. Where possible, the researcher used tasks or assessment items previously developed by people assumed to be experts at statistics task design as the basis for the tasks. This included assessment tasks published by NZQA and tasks used in peer-reviewed statistics education research. Care was also taken with the wording of tasks to minimise interpretation issues, in particular ambiguity. The tasks developed were also checked by the researcher’s supervisor and one other statistics educator before they were used with participants in the research. Across the six tasks designed by the researcher, care was taken to use different questions in an attempt to measure the same understanding. For example, in an attempt to validly measure a teacher’s understanding of shape, Task C required participants to physically sketch the shape of a distribution, Task D required participants to discuss a hypothetical student’s understanding of shape, and Task E required participants to explain the criteria they used to sketch the shape of a distribution. In this way, more than one measure was used and compared. As this research aimed to explore teachers’ understanding of teaching probability distribution modelling, as assessed by AS91586, the researcher also ensured that participants in the study did in fact have experience teaching this material.

A structured task-based self-interview was used as the instrument to collect data for the research so that there were similar conditions for the participants within the study. Although the results of this research cannot be extended beyond the participants in this small exploratory study, the methods used within the research are considered to be reliable if used to replicate this research with another group of New Zealand teachers with experience teaching AS91586. Thematic analysis
was used to analyse the task responses and the researcher took care to document the decisions made throughout this analysis process.

Triangulation is the combining of methods and data, using both quantitative and qualitative approaches, and adds credibility to the study by strengthening the conclusions made (Patton, 2015). To further strengthen the validity and reliability of this research, triangulation was used in two ways. Firstly, in Chapter 2 the analysis of documents published by teachers experienced with teaching AS91586 was used as another source of data concerning teacher knowledge of probability distribution modelling. This allowed the researcher to compare the results of the research to evidence collected from external sources. Secondly, triangulation was used within the design of the study with the embedding of a randomised experiment within the qualitative online interview. Hence key inferential interpretations made by the researcher within the qualitative analysis could be validated against the results of the randomised experiment.

Ethics
Permission to conduct the study was granted by the University of Auckland Human Participants Ethics Committee (see Appendix B). Written and informed consent was required before each teacher could participate in the research, and all participants had the right to withdraw from the study prior to a predetermined date. All data were reported in such a way that the participants and their respective schools cannot be identified. Each potential participant in this study needed to be screened by the researcher to ensure they were a legitimate teacher with experience teaching AS91586, so participation in the research project was not anonymous. To preserve anonymity in the teacher responses to the online tasks, the login details used to access the secure site hosting the online tasks were system generated and known only to the participant. In this way the researcher was not able to associate teacher responses with individual teachers or schools. Because the researcher was aware of which teachers had volunteered to participate in the study, the researcher gave assurance that she would maintain a professional relationship with each participant. The researcher also gave assurance that the participant’s decision to participate or not would not affect any
relationship the participant may have had with the researcher. Teachers completed the online tasks in their own time and at a time and place that suited them. To help resolve the issue of additional work for teachers during term time, the data collection period included part of the school holidays.

**Researcher bias**

The researcher has been involved with the development of curriculum and assessment materials for the New Zealand Curriculum and associated NCEA Achievement Standards. This work required decisions to be made about teaching and assessment related to probability distribution modelling. In particular, the researcher contributed to the material developed within New Zealand to support the teaching of probability distribution modelling (AS91586). To minimise this bias, the researcher took an objective stance on her own work, and included critique of her own documents alongside the critique of others (see Chapter 2, Martin, 2014). As thematic analysis does not happen in a vacuum and tends to be driven by the researcher's theoretical interests (Braun & Clarke, 2006), there was the need for the researcher to actively seek data and responses that contradicted or challenged her pre-conceived views on probability distribution modelling.
Chapter 5

Results

5.1 Introduction

The responses to each task were analysed using the methods outlined in Chapter 4. This chapter presents a task completion analysis and the key results emerging from each task, Tasks B to F (see Appendix A).

5.2 Task completion analysis

The task-based interview completed by teachers consisted of six sequential tasks, Tasks A to F. Some teachers ceased their participation in the study before completing all tasks. Therefore, the results presented from each task vary in terms of the numbers of teachers in the study who participated. Nine of the 17 teachers in this study completed all tasks. A comparison was made of the background of the teachers who did or did not complete all tasks. Nonsignificant differences were found between the two groups in terms of the proportion of females, mean years teaching overall and mean years teaching AS91586. A notable feature is that among the 9 teachers who completed all tasks were the only four teachers within the study who completed a Statistics major or equivalent for their tertiary qualification.

5.3 Results from Task B: Preparing to teach AS91586

For Task B, teachers were asked questions related to the preparation of lessons for teaching AS91596. All 17 teachers in the study completed this task.

5.3.1 Key understandings described by teachers

In preparing lessons for teaching AS91586, 16 of the 17 teachers who completed Task B used teacher-published material from the Census at School New Zealand website, with 14 of these 16 teachers also making use of NZQA assessment documents. Eleven of the 17 teachers used material from the Senior Secondary Guide as part of their preparation, a lower number than those who used school-based materials (13 teachers) or textbooks (12 teachers). None of the teachers
who completed all tasks in this study had used the Senior Secondary Guide document *Lateness: Choice or chance*.

When asked to describe what could be considered three key concepts, ideas or understandings for a sequence of lessons towards AS91586, 14 of the 17 teachers who completed this task gave substantial descriptions. The three descriptions given by each teacher were first separated and the descriptions across all teachers were then sorted into groups based on the similarity of the understanding described. During this initial process, no framework was actively used to judge similarity, with groups created initially based on comparisons of aspects of descriptions, such as terms used.

The groups of descriptions were reviewed and in some cases merged, based on the researcher’s view that some descriptions appeared to be describing understandings that were specific to different steps within a probability modelling process while others appeared to be describing conceptual understandings. At this stage of the analysis, for some teachers two descriptions were combined to create one key understanding or other descriptions were separated to create two key understandings. This was done to ensure consistency within the group and equity in scope between groups and resulted in a final set of 45 descriptions across eight distinct groups. Table 7 shows the results of this analysis, which suggest that these teachers prioritised theoretical probability distributions or theory-driven modelling as key understandings necessary for learning AS91586.

Teachers were also asked to describe two aspects they found students struggled with when learning about probability distribution modelling. Thirteen of the 17 teachers who completed this task gave complete descriptions. Twenty-one of these 26 descriptions concerned theory-driven probability distribution modelling. That is, the application of a theoretical probability distribution to solve a word problem.
<table>
<thead>
<tr>
<th>Understanding described</th>
<th>Number of descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modelling (concept)</td>
<td>9</td>
</tr>
<tr>
<td>Understanding models are not the same as reality but can be</td>
<td></td>
</tr>
<tr>
<td>used to approximate a real situation; Understanding that a</td>
<td></td>
</tr>
<tr>
<td>probability distribution can model the relative likelihood</td>
<td></td>
</tr>
<tr>
<td>of a set of numerical outcomes; Using models as a way to</td>
<td></td>
</tr>
<tr>
<td>quantify uncertainty</td>
<td></td>
</tr>
<tr>
<td>Theoretical probability distributions</td>
<td>8</td>
</tr>
<tr>
<td>Recalling conditions; Recalling properties, including</td>
<td></td>
</tr>
<tr>
<td>features such as shape; Understanding effect of parameter</td>
<td></td>
</tr>
<tr>
<td>change; Area under the curve = 1; Discrete versus</td>
<td></td>
</tr>
<tr>
<td>continuous</td>
<td></td>
</tr>
<tr>
<td>Selection of theoretical probability distribution model</td>
<td>6</td>
</tr>
<tr>
<td>based on written description of situation</td>
<td></td>
</tr>
<tr>
<td>Comprehending written problems; Identifying features of</td>
<td></td>
</tr>
<tr>
<td>contextual description; Matching features of context with</td>
<td></td>
</tr>
<tr>
<td>features of theoretical probability distribution</td>
<td></td>
</tr>
<tr>
<td>Fit of probability distribution model to sample data</td>
<td>5</td>
</tr>
<tr>
<td>Using probabilities to determine which model fits best;</td>
<td></td>
</tr>
<tr>
<td>Using visual displays; Taking into account variation</td>
<td></td>
</tr>
<tr>
<td>related to sample size; Using sample-to-population</td>
<td></td>
</tr>
<tr>
<td>generalisations</td>
<td></td>
</tr>
<tr>
<td>Application of probability distributions and discrete</td>
<td>5</td>
</tr>
<tr>
<td>random variables</td>
<td></td>
</tr>
<tr>
<td>Use of probability distributions to calculate probabilities;</td>
<td></td>
</tr>
<tr>
<td>Interpreting word problems; Calculating mean and standard</td>
<td></td>
</tr>
<tr>
<td>deviation; Performing inverse calculations; Using</td>
<td></td>
</tr>
<tr>
<td>continuity corrections; Combinations and transformations of</td>
<td></td>
</tr>
<tr>
<td>discrete random variables</td>
<td></td>
</tr>
<tr>
<td>Distribution (concept)</td>
<td>5</td>
</tr>
<tr>
<td>Understanding the term distribution in terms of variation;</td>
<td></td>
</tr>
<tr>
<td>Linking outcomes and probabilities to statistics such as</td>
<td></td>
</tr>
<tr>
<td>$E(X)$ and $SD(X)$; Describing shape of distribution</td>
<td></td>
</tr>
<tr>
<td>Randomness (concept)</td>
<td>4</td>
</tr>
<tr>
<td>Understanding randomness and random variables; Uncertainty/</td>
<td></td>
</tr>
<tr>
<td>probabilistic versus deterministic</td>
<td></td>
</tr>
<tr>
<td>Different ways of thinking about probability</td>
<td>3</td>
</tr>
<tr>
<td>The difference between distribution of true probabilities,</td>
<td></td>
</tr>
<tr>
<td>distribution of model estimates of probabilities and the</td>
<td></td>
</tr>
<tr>
<td>distribution of experimental estimates of probabilities</td>
<td></td>
</tr>
</tbody>
</table>
The key areas of difficulty for theory-driven probability distribution modelling identified by teachers were:

- Understanding the question and not being misled by words
- Remembering the key aspects of each probability distribution
- Selecting the probability distribution needed to solve the problem
- Mathematical skills such as inverse calculations, expectation algebra, and continuity corrections
- Interpreting probability language

Only four of the remaining five descriptions concerned student difficulty with probability distribution modelling involving sample (experimental) data. This is perhaps not surprising given the large number of key understandings described that are of a theoretical nature.

5.3.2 Understanding of the three types of probability distributions

Teachers were asked to describe a learning activity that they had used with students to build understanding of distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities. Teachers were also asked to specifically explain how the activity they had described helped build understanding of these different ways to think about probability distributions. Twelve of the 17 teachers in this study gave complete descriptions.

Four of the 12 teachers’ descriptions used examples or past exam questions for AS91586 to demonstrate the three ways of thinking about probability distributions. The remaining eight teachers described using hands-on activities requiring the students to collect or generate data from a real situation. Situations described included:

- Tossing plastic butterflies and observing how many land on their feet
- Randomly selecting multilink cubes from a bucket and counting how many are a certain colour
- Counting the number of pieces of hokey pokey per scoop of ice-cream.
A common thread for the hands-on activities described was that the situations were ones where the true probability distribution was not known, but for which a theoretical probability distribution from those expected to be known at curriculum level eight (e.g. normal, Poisson, binomial, triangular, uniform) could be considered. Teaching strategies described within these activities included:

- Asking students to sketch what they thought could be the shape of the true probability distribution before collecting any data
- Using sampling type situations where the object is destroyed in the measurement process and so the development of a model would be apparent to students, e.g. how long batteries last for
- Representing the data collected in graphs so visual aspects of the distributions could be described and compared
- Comparing distributions for data collected or generated between students to see sampling variation
- Considering other sources of variation in the data by using contextual knowledge about the situation being modelled.

The learning activities described appear to be consistent with the definitions of distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities discussed in Chapter 2, indicating that the teachers in this study have a secure understanding of each these three types of probability distribution.

5.3.3 Summary

Teachers were able to explain the differences between the distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities and provided some good examples of teaching activities that would support the understanding of these terms. Teachers considerations about the key understandings required for the learning AS91586 were mainly theory-driven, as were their considerations of difficulties faced by students.
5.4 Results from Task C: Sketching shapes of distributions

For Task C, teachers were asked to sketch the shape of six different distribution using either a touchscreen or computer mouse. Twelve of the 17 teachers in the study completed this task.

5.4.1 Similarity of the shape sketches drawn

It was conjectured that if teachers shared a common understanding of how to sketch the shape of a distribution then their overlaid plots for each distribution would reveal a blurry but distinct shape. That is, although the placements of each shape sketch might vary from teacher to teacher for each distribution due to the use of a computer mouse or touchscreen and to each teacher's skill in using these devices, when the shape sketches were overlaid on the sample plot, they should be similar enough to collectively “paint a picture” of the reference distribution. To explore this conjecture, for each of the six distributions, the shape sketches from all 12 teachers were combined and overlaid on the same plot without the reference distribution, as shown in Figure 25.

(a) Distribution 1          (b) Distribution 2          (c) Distribution 3

(d) Distribution 4          (e) Distribution 5          (f) Distribution 6

*Figure 25: Task C shape sketches for each distribution produced by combining the sketches from all 12 teachers and overlaid on the same plot without the reference distribution*
Based on visual inspection, Distribution 1 displayed the most similar shape sketches and Distribution 4 displayed the most dissimilar shape sketches. A notable difference within the overlaid shape sketches for some of the distributions is in the number of distinct “bumps” drawn or “steepness” of curves. As the overlaying process obscured some features of the individual shape sketches, the shape sketches for each distribution were reviewed individually and grouped by similarity in terms of “bumps” and “steepness”, as shown in Table 8.

The distributions have been vertically displayed in Table 8 in order of smallest sample size to largest sample size. The grouping of individual shape sketches for each distribution in terms of “bumps” and “steepness” indicates how specific the shape sketch was. This arrangement reveals a possible relationship between the sample size of the distribution and the shape sketched. Generally, the fewer data that were displayed in the distribution, the less similar the shape sketches were with each other, using the number of groups of sketches created as a measure of similarity. The high level of consistency in sketching unimodal shapes for Distributions 1 and 5 may be partially influenced by the fact they are also both positively skewed and more easily recognised as a shape.

Although Distributions 1 and 3 both have features which indicate potential bimodality, Distribution 1 represents a large amount of data and Distribution 3 represents a smaller amount of data. However, none of the teachers indicated bimodality in their shape sketches for Distribution 1, in contrast to Distribution 3 where eight of the twelve teachers indicated bimodality in their shape sketches. Taking into account the results for Distribution 2, where some teachers also indicated bimodality in their shape sketches, it appears that teachers were more likely to sketch more specific shapes for smaller samples, even though features of distributions of small samples can be artefacts of noise. Distribution 2, for example, shows only four teachers who appear to take the view that the features of the sample distribution are noise, and so sketch a uniform distribution shape.

Note: The visual inspection was done using overlayed plots viewed digitally in colour and with a transparency feature. Individual sketches within the overlayed plots are unable to be seen in black and white hardcopy printed form.
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Graph of distribution</th>
<th>Less specific</th>
<th>More specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Potentially unimodal</td>
<td>1 teacher</td>
<td>8 teachers</td>
</tr>
<tr>
<td>2</td>
<td>Potentially bimodal</td>
<td>4 teachers</td>
<td>3 teachers</td>
</tr>
<tr>
<td>6</td>
<td>Potentially unimodal</td>
<td>2 teachers</td>
<td>5 teachers</td>
</tr>
<tr>
<td>3</td>
<td>Potentially bimodal</td>
<td>4 teachers</td>
<td>8 teachers</td>
</tr>
<tr>
<td>5</td>
<td>Potentially unimodal</td>
<td>11 teachers</td>
<td>1 teacher</td>
</tr>
<tr>
<td>1</td>
<td>Potentially bimodal</td>
<td>12 teachers</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Comparison of shape sketches produced for Task C
5.4.2 Further categorising shape sketches

When each shape sketch was viewed individually with reference to the distribution displayed, further characteristics of the shape sketch were able to be explored by linking features of the shape sketched to features of the reference distribution. Each shape sketch was initially intended to be scored using the criteria developed by Arnold (2013) in her research with students describing distributions (see Table 9).

<table>
<thead>
<tr>
<th>Score</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No shape drawn</td>
</tr>
<tr>
<td>1</td>
<td>Have redrawn the graph with the points again</td>
</tr>
<tr>
<td>2</td>
<td>Have drawn a curve with multiple bumps; basically drawing the city skyline; drawing literally around the data</td>
</tr>
<tr>
<td>3</td>
<td>Smoother curve, but additional bumps OR connected the dots to make a polygon shape rather than a smooth curve</td>
</tr>
<tr>
<td>4</td>
<td>Smooth representative curve</td>
</tr>
</tbody>
</table>

However, applying these criteria to the sketches made by teachers in this study proved problematic, particularly in terms of deciding whether a curve was “representative”. Each of three distributions used in Arnold’s assessment of shape sketches (2013, p. 220) represented a large amount of data with fairly strong signals in terms of skewness and modality. In this respect, it may have been easier for Arnold to judge if a shape sketch was representative. Most shape sketches made by teachers in this study were smooth curves, and Arnold’s criteria did not capture the visual differences the researcher could see in the sketches. As the study was interested in how sketches of shape may differ when small sample distributions are used, Arnold’s criteria needed to be adapted to allow for the impact of this factor. Additionally, there was a need to take into account the nature of the distributions within the context of probability modelling. For example, the triangular and uniform distributions are polygon shapes, and so shape sketches
based on these models may not represent a lower level of understanding regarding shape. The researcher also wanted to categorise shape sketches based on their vertical placement with respect to the distributions, for example, whether the teacher sketched *above* the distribution or *through* the distribution.

Drawing on some of the features of the Distribution Description Framework developed by Arnold (2013, p. 217), criteria were developed to evaluate sketches based on three characteristics: smoothness, vertical placement, and horizontal placement. These characteristics can be used to categorise a teacher’s general approach to sketching across different distributions presented to them. Due to the use of a mouse and/or touch screen to sketch the shape, the researcher was unable to apply the criteria for *smoothness* as intentional “bumps” could not be distinguished from “bumps” caused by the use of a digital tool to capture the shape. The *placement* criteria were used to categorise all of the sketches made by each teacher, and the dominant characteristic of each teacher was determined based on which characteristic each teacher used the most often across each of their six shape sketches (see Table 10).

Common characteristics of the shape sketches made across all distributions were that the sketch was above the distribution and the horizontal position the “peak” or “bump” of the shape matched the outcome with the highest frequency in the distribution. Ten of the 12 teachers who completed this task predominantly sketched shapes of distributions that displayed both of these characteristics. This suggests a relationship between sketching above a distribution and the highest frequency observed, an outline view, which does not appear to take into account sample size. Such a relationship raises a question about why, for example, the highest frequency outcome influences the shape but not the lower frequency outcomes. The two teachers who generally sketched through the distributions did not line up the “peak” of their shape sketch with the individual outcome with the highest frequency, suggesting more of a “smoothing” approach to the sketching of the shape.
Table 10: Criteria for sketching the shape of the distribution of a variable for general characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Category, criteria and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothness</td>
<td>Consistent: Predominantly smooth curve with at most two small bumps</td>
</tr>
<tr>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>Placement (vertical)</td>
<td>Above: Shape is predominantly drawn above the distribution, lots of white gaps below shape</td>
</tr>
<tr>
<td></td>
<td>10 teachers</td>
</tr>
<tr>
<td>Placement (horizontal)</td>
<td>Highest frequency: The peak(s) of the shape (or position of the horizontal line) match the outcome(s) with the highest frequency</td>
</tr>
<tr>
<td></td>
<td>10 teachers</td>
</tr>
</tbody>
</table>

5.4.3 Summary

The shape sketches were generally more consistent between teachers for distributions when $n$ is large than for distributions when $n$ is small and it appears that shape sketches are also more specific for smaller distributions. A tendency for teachers to sketch the shape over the top of a distribution was observed, a method which could facilitate the sketching of specific features, particularly “peaks” or “bumps.”
5.5 Results from Task D: Assessing student understanding

For Task D, teachers were asked to identify the misunderstanding(s) demonstrated by two students, A and B, regarding the appropriateness of a probability distribution model. Twelve of the 17 teachers in the study completed this task. These results focus on the responses related to Student A, who had concluded that a normal distribution was not a good model because the shapes of the sample distribution and the model probability distribution were not similar. In particular, Student A had described the sample distribution with sample size 50 as negatively skewed. The teachers first discussed the nature of the misunderstanding of Student A, namely describing the shape of small samples. They then discussed what activities could be used to address any misunderstanding(s), namely through the use of simulation.

5.5.1 Describing the shape of small samples

It was intended by the researcher that the key misunderstanding(s) demonstrated by Student A was to infer the shape of the true probability distribution from one small sample of 50 and to reject the proposed model (a symmetric distribution) because its shape was different from the sample distribution (a negatively skewed distribution). Seven of the 12 teachers who completed this task discussed these specific misunderstandings, making comments such as:

...possibly insufficient sample size to make those comments about comparing shape... [Teacher 17]

...this student assumes that the skewness that is identifiable in the graph is a feature of the parent distribution ... [Teacher 3]

However, the other five teachers did not agree with the Student A’s description of the sample distribution as negatively skewed (see Figure 26b). A common aspect of the responses given by teachers who did not agree with the description of the sample distribution as negatively skewed was discussion of the lowest jellybean length. For example, Teacher 13 stated that “...The amount of skew is fairly minimal and is exaggerated by a single data point (approx. 16.4)...”. This view is reinforced when four of these five teachers suggest that Student A should ignore the smallest value, and do the following instead:
...look at where the majority of the jelly beans are and sketch the shape based on the majority of the jelly beans... [Teacher 15]

...sketch the outline with/without the min. size and compare the shape... [Teacher 16]

The researcher has attempted to demonstrate the thinking of these five teachers through the sketch made in Figure 26c. It should be noted that the sample distribution is real data, and the minimum value (jellybean length) is a real outcome from the true probability distribution.

![Figure 26: Example of sketches of the sample distribution used in Task D](image)

5.5.2 Use of simulation

Nine of the 12 teachers who completed Task D provided detailed descriptions of activities they would use to help address the misunderstanding(s) demonstrated by Student A. Seven of these nine teachers specifically discussed the use of simulation in their descriptions. Two themes were identified across these descriptions: the importance of visualising sampling variation and the infrequent use of probability modelling language.
The importance of visualising sampling variation. Many teachers shared a belief that using a simulation would allow Student A to see why they were incorrect to state that “... the normal distribution is not a good model for the lengths of jelly beans produced by the confectionery company ...”, a statement that Student A had made based on comparing the shapes of the sample distribution and the model distribution. In communicating this belief, it appears that teachers assumed that Student A had not had enough exposure to visualising sampling variation, making statements such as:

... Student A does not have an appreciation of how much random samples of size 50 can differ between samples, what is likely and unlikely shapes to get... [Teacher 1]

... Student A is not aware that a sample of 50 from a normally distributed population will not look "normal"... [Teacher 9]

A common strategy suggested was for teachers to show students through multiple simulated samples of the same size as the original sample that the shape of the sample distribution would vary, for example:

... I may use a sampler/simulator of some description and demonstrate that further samples of 50 will look different to each other... [Teacher 2]

.... Set up a system for showing multiple samples of jellybeans from the same population to show that we wouldn't necessarily see the same shapes in each sample... [Teacher 13]

While the need for students to be able to visualise variation was communicated by most teachers, doubt was also expressed about whether the use of visualisations was effective. For example, Teacher 4 stated that they were unsure “... if the students are focusing on what I want them to understand when I show them many different iterations of a simulation ...”.

The infrequent use of probability modelling language. Throughout the nine descriptions of activities proposed to help build Student A’s understanding, there
was only one use of the word *model* and no use of the word *true*. The minimal use of these words was noted by the researcher as surprising for three reasons:

1. The problem presented was about whether a proposed normal distribution model was a *good model*, so the word model was in the question stem.

2. The teachers frequently discussed the use of simulations as part of their descriptions, yet simulations require a model.

3. The *true probability distribution* is unknown, an understanding that was well communicated by these teachers in Task B.

Instead, language associated with *sample-to-population* inference was more frequent. For example, data that would be generated through simulation from a *model distribution* was invariably referred to as generating *sample data or samples* from a *population*, as shown in the excerpts below:

... I would make an excel simulation of random samples of 50 from a normally distributed population ... [Teacher 4]

... I would give the student samples from a normal distribution ... [Teacher 14]

While Teacher 9 gave an alternative approach that appeared to avoid the use of simulation - setting up a “fake” normally distributed population of 1000 jelly beans, and then getting students to use software to sample without replacement from this population – this population of 1000 jelly beans would itself not be a true normal distribution, rather an approximate empirical distribution. Because a normal distribution is a theoretical mathematical model, the generation of “samples” from this distribution requires the use of simulation. Also, a probability distribution model is not a *population* of finite objects.

Therefore, for Task D teachers would need to conceive the lengths of jellybeans produced by the confectionary company as a *large finite population*, in order to use repeated sampling to demonstrate to students that the observed sample could *come from* a “normally distributed population”. This conceptualisation can be seen
in the description made by Teachers 2 and 14, who identify the misunderstanding demonstrated by Student A as:

... sample to population inference directly based just on smallish sample...
[Teacher 2]

...missing the sample population link... [Teacher 14]

While the use of the words *sample* and *population* are not in themselves an indication of misunderstanding, it was not clear to the researcher that the teachers viewed the normal distribution presented in the problem as a possible *model* for the lengths of jelly beans produced by the confectionary company. That is, the use of sampling language may not clearly signal the idea of *modelling* using probability distributions.

5.5.3 Summary

Through the discussion of Student A’s misunderstandings concerning the shape of the small sample distribution, it became apparent that some teachers view shape as 'descriptive' (the shape of the sample distribution they can see) while others view shape as 'inferential' (the shape of the underlying probability distribution that is not seen). Although teachers described using simulations to build ideas necessary for probability distribution modelling, it was not clear that they conceived a probability distribution as a model rather they conceived it as a population or the true probability distribution.
5.6 Results from Task E: Capturing modelling processes (the randomised experiment)

Ten of the 17 teachers in the study completed this task. Teachers who completed this task were unaware that they had been randomly assigned to receive one of two different versions of this task. The only difference between the two versions of this task, and the treatment variable for this experiment, was the amount of data supplied (the sample size).

A comparison was made between the background of teachers within each treatment group of the randomised experiment. Nonsignificant differences were found between the two treatment groups in terms of the proportion of females, mean years teaching overall and mean years teaching AS91586. All four of the teachers with statistics majors or equivalent for their tertiary qualification completed Version 2 of Task E. It was conjectured that there would be no overall difference between the two groups of teachers when reviewing responses to questions within the task.

Before being asked to sketch the shape of the sample distribution and then test the visual fit of the proposed model, teachers were asked to describe the process they would use to investigate whether the number of emails received per hour to an email account could be modelled by a Poisson distribution. Sample data were provided from the situation and teachers were asked to describe any expectations, calculations, tests, drawings, graphs, models or theoretical probability distributions, conditions, assumptions or limitations that they would use or consider as part of their investigation. No guidance was given to teachers as to how many steps to describe, although there were only 10 boxes available in the task.

5.6.1 Framework analysis of modelling processes described

An initial review of the steps described by teachers who completed this task revealed that with respect to the nature of the process described, the amount of data available to be used was not an influencing factor. Specifically, only three of the 10 teachers discussed the number of hours of email data ($n = 48$ or $n = 480$).
Therefore, the processes described by each teacher were combined from both treatment groups and analysed together.

It was anticipated by the researcher that the steps described by the teachers could be synthesised by extracting and combining the key aspects of each investigative process. Therefore, the first attempt at analysis involved reading through each set of steps described and attempting to summarise the investigative process described in terms of key phases and the sequence of these phases, for example, something similar to the data enquiry cycle, also known as PPDAC (problem, plan, data, analysis, conclusion). However, summarising the investigative process described in terms of a sequence of phases proved problematic. While the investigative process was described in ordered steps, each step often made connections between contextual and statistical knowledge and/or combined theory-driven and data-driven strategies. It became apparent during the analysis that the order of the steps described was not as informative as what probability modelling knowledge teachers integrated throughout their investigative process.

For example, during the first attempt at analysis it was identified that the first step described by teachers could be classified as being either theory-driven or data-driven. Five of the 10 teachers described checking if the conditions for the distribution were met based on the information presented, a theoretical approach based on comparing contextual knowledge and knowledge of the specific probability distribution. For the remaining five teachers, the first step described involved creating a graph of the sample data and describing its features, or calculating the mean of the sample data first without any initial visualisation of the data.

However, rather than the first step description giving primacy to the remaining step descriptions in terms of overall approach (theory-driven versus data-driven), analysis of the remaining step descriptions found that most teachers went on to use both theoretical and data approaches regardless of the nature of the first step described.
Therefore, it was decided instead to use framework analysis to analyse the investigative process described as a whole. Specifically, rather than focusing on a linear process, it was decided to encode the investigative process described in a framework that demonstrated what knowledge was described and how this knowledge was linked. The framework used for this analysis was an adapted version of the proposed statistical modelling framework.

In this adapted version of the statistical modelling framework, the four components of the framework are based on knowledge:

- Contextual knowledge: Emails arriving to an email account every hour and any conjectures or knowledge about this situation, including whether the process is random
- Theoretical probability distribution model knowledge: Features and conditions of the Poisson distribution
- Data knowledge: Features of the sample data provided
- Sampling variation knowledge: Features of the simulated data

The steps described by each teacher were then reviewed to identify which knowledge components were utilised, what specific knowledge was used, and how these knowledge components were connected and integrated in the teacher’s reasoning process. The data were then encoded onto one statistical modelling framework using an evidence-based synthesis of the knowledge and reasoning described. In cases where incorrect knowledge was stated by the teacher (for example, that for a Poisson distribution that the mean = standard deviation and that lambda must be small) this was ignored and not used by the researcher in the analysis. When comments were made by teachers that were insightful but not able to be encoded on the framework, these were noted and presented supplementary to the analysis. The resulting encoded statistical modelling framework is shown in Figure 27.
It was found that there were two main connections made by teachers between knowledge components:

1. Contextual knowledge and theoretical probability distribution model knowledge (comparing conditions of model with situation)

2. Data knowledge and theoretical probability distribution model knowledge (comparing features of distributions, including comparing theoretical probabilities with observed proportions)

As noted earlier, seven of the 10 teachers did not discuss the number of hours of email data ($n = 48$ or $n = 480$). The lack of attention to sample size seems to be problematic for connection 2, as it requires the comparison of sample data with the model and these comparisons should consider how much data are available.
The framework analysis (Figure 27) highlights the lack of connection made between the data knowledge component and the contextual knowledge component. This connection requires inferential reasoning. However, despite teachers giving consistent and clearly articulated descriptions of *distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities* for Task B (discussed in 5.3.2), there was no mention of the *distribution of true probabilities* or use of the word *true* by teachers in the descriptions of modelling steps.

In terms of how teachers used both connections 1 and 2 in their reasoning about the suitability of the probability distribution model, three approaches emerged. Using one approach, two teachers believed that the conditions for the Poisson distribution should be checked first, and if these did not hold, then the investigation should not proceed. Using another approach, one teacher also considered the conditions for the Poisson distribution, but was happy for not all of them to be met if the Poisson distribution model looked like it was a good fit to the sample data. The final and most common approach for teachers who had made links between contextual knowledge and the conditions of the Poisson distribution was to discuss any issues identified as limitations to the model.

Two of the 10 teachers, in addition to discussing sample size, described using simulated data as part of the probability distribution model fitting process. This connection can be summarised as follows:

3. Theoretical probability model knowledge and sampling variation knowledge (comparing features of distributions, including comparing theoretical probabilities with simulated proportions)

Note that this connection does not directly compare data knowledge with the sampling variation knowledge, therefore there is no connection made between the data knowledge component and the sampling variation knowledge component. The framework analysis (Figure 27) shows that the probability distribution model is the centre of teachers’ thinking when probability distribution modelling, as all three connections involve the theoretical probability model knowledge component.
5.6.2 Criteria described for sketching shapes of distributions

In Task E, teachers were also asked to sketch the shape of the distribution of number of emails received per hour. Note five teachers were randomly assigned Version 1 \((n = 48)\), the other five Version 2 \((n = 480)\). Using the shape criteria developed in Table 10, three teachers’ shape sketches for each version were categorised as being placed above the distribution. All five of the shape sketches made for Version 1 of the task displayed “peak(s)” or “bump(s)” that matched the outcome(s) with the highest frequency in the distribution, and three of the five shape sketches met this same criterion for Version 2. Therefore, it appears the shape sketches made for each version of this task are similar based on these criteria, despite the sample sizes for each version of this task being quite different. The small number of teacher sketches used for this analysis, however, makes this conclusion of similar shape sketches very tentative.

The similarities between shape sketches made for each different version of this task are even more apparent when the shape sketches for each of the 10 teachers are arranged by matching pairs of visually similar shape sketches between Versions 1 and 2 (see Table 11). The arrangement of shape sketches in Table 11 also shows two teachers from both versions sketched a bimodal distributional shape. This is a surprisingly low number for Version 2 of Task E, because the bimodality visible in the distribution should be interpreted as a very strong signal for the distribution, as the counts associated with outcomes 0 and 2 would be over 100 each from a total sample size of 480.

The teachers in this task were also asked to describe the criteria they used to sketch the shape of the distribution. To explore further the reasons why the shape sketches were quite different when comparing sketches from the same version of the task, the criteria described by teachers were analysed to identify the different shape sketching methods used. Nine of the ten teachers who completed the task gave complete descriptions of the criteria used to sketch the shape of the distribution. These criteria described were categorised, and the results are presented in Table 12.
Table 11: Shape sketches for Task E, matched by pairs of visually similar sketches across Versions 1 and 2 of the task

<table>
<thead>
<tr>
<th>Version 1 (n = 48)</th>
<th>Version 2 (n = 480)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unimodal</strong></td>
<td></td>
</tr>
<tr>
<td>Teacher 1</td>
<td>Teacher 2</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>Teacher 15</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>Teacher 16</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>Teacher 12</td>
</tr>
<tr>
<td><strong>Bimodal</strong></td>
<td></td>
</tr>
<tr>
<td>Teacher 5</td>
<td>Teacher 9</td>
</tr>
<tr>
<td>Teacher 6</td>
<td>Teacher 13</td>
</tr>
<tr>
<td>Teacher 7</td>
<td>Teacher 3</td>
</tr>
<tr>
<td>Teacher 8</td>
<td>Teacher 1</td>
</tr>
<tr>
<td>Teacher 9</td>
<td>Teacher 13</td>
</tr>
<tr>
<td>Teacher 10</td>
<td>Teacher 5</td>
</tr>
</tbody>
</table>

Number of emails received per hour
Table 12: Categories of criteria described to sketch the shape of a distribution

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
<th>Example excerpt</th>
<th>Number of teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outline</td>
<td>The sketch should be a smooth outline of the data distribution, with the peak(s) or highest point(s) determined by the outcome(s) with highest frequency</td>
<td>“…imagining how a heavy blanket would drape itself over the columns…” [Teacher 13]</td>
<td>3</td>
</tr>
<tr>
<td>Context</td>
<td>The sketch should be informed by contextual knowledge of the variable, use context to decide which features of data distribution to incorporate in shape</td>
<td>“…it will be bimodal because there are a lot of hours in the middle of the night when no-one sends me an email…” [Teacher 3]</td>
<td>3</td>
</tr>
<tr>
<td>Levelling</td>
<td>The sketch should try to balance out the heights of each bar or stack of dots, to take into account sampling variability or noise</td>
<td>“…I thought that the up and down at 1 to 4 was due to sampling variability, so I didn’t follow that exactly…” [Teacher 4]</td>
<td>2</td>
</tr>
<tr>
<td>Model</td>
<td>The sketch should be based on the model distribution, use model to decide which features of data distribution to incorporate in shape</td>
<td>“…Assumed a Poisson with mean around 1 or 2…” [Teacher 2]</td>
<td>1</td>
</tr>
</tbody>
</table>

Interestingly, seven of the nine descriptions use knowledge beyond what could be seen in the distribution. Teachers using the context criterion attempted to use knowledge about the behaviour of emails to the email account, teachers using the model criterion attempted to use knowledge about the proposed model for number of emails arriving per hour, and the teachers using the levelling criterion attempted to use knowledge about how much variation there would be between different samples of email data. None of the nine teachers made reference to how much data was represented in the distribution and all of the teachers drew sketches that were based on all outcomes in the distribution. For teachers who completed Version 1 of this task, their actions are different from what they suggested they would do in Task D, and that was to take sample size into account.

5.6.3 Informally testing the fit of a probability distribution model

Teachers were also asked in Task E to discuss the appropriateness of a specific Poisson model in terms of the visual fit of the model to the sample data provided. All ten teachers who completed this gave responses that demonstrated what
method they used to informally test the fit of the probability distribution model. Examples of good responses that the researcher expected for each version of the task are shown in Figure 28.

**Example of good response:** This Poisson distribution model is appropriate (it could not be ruled out as an appropriate model), as when taking into account the small amount of data available \( (n = 48) \), the differences seen visually between the model (expected) and sample (observed) counts could be explained by sampling variation.

**Example of good response:** This Poisson distribution model is not appropriate, as when taking into account the large amount of data available \( (n = 480) \), the differences seen visually between the model (expected) and sample (observed) counts could not be explained by sampling variation.

*Figure 28: The two versions of Task E, presented with examples of good responses*

The response given by each teacher was recorded as either concluding that the proposed model was appropriate or not appropriate. If there was not a clear overall statement, then the response was coded as “not appropriate”, for example, the response given below:

In terms of a visual fit of the model to the sample data, the model underestimates the number of hours where no emails are received (and somewhat for the number of hours where 4 emails have been received), and overestimates then number of hours where 1 email is received [Teacher 1]

Despite the fact that the two versions of this task had been designed to elicit different responses, three Version 1 teachers responded the same way as three Version 2 teachers, by discussing the Poisson distribution model as being
appropriate. Therefore, there is no evidence that sample size affected these teachers’ decision when “making a call” about the fit of the model. Only three of the ten teachers who completed Task E made reference to the sample size \( (n = 48 \text{ or } n = 480) \) when considering the fit of the proposed model: two Version 1 teachers and one Version 2 teacher.

It should be noted, however, that three of the six teachers who concluded the proposed model was an appropriate model either expressed a change of mind or stated a lack of confidence in their conclusion within their response, for example:

... so it [the model] probably isn’t that appropriate... [Teacher 5]

.....things that make me less confident are.... [Teacher 13]

Therefore, in addition to sample size not being considered by all of the teachers when *informally testing the fit of a probability distribution*, the displayed reluctance to commit to a conclusion about the appropriateness of a model further suggests there are issues with the criteria used by teachers. Consequently, reasons given in support of the conclusions made were reviewed to identify the criteria used by teachers to decide on the visual fit of the model. It was found that teachers used six criteria for *informally testing the fit* of the probability distribution model used in this task, with nine of the ten teachers using more than one criterion (median number of criteria used = 3). These criteria are presented in Table 13.

The two most used criteria to compare the proposed model with the sample data – *individual outcomes* and *shape* – were used by teachers irrespective of which version of the task was completed. Five teachers used both the *individual outcomes* and *shape* criteria as part of their reasoning. It is not surprising that these two criteria were used most frequently as they are related: the shapes of the model and sample distributions are based on the individual outcomes within each distribution. In fact, if the shape of the sample distribution was sketched by tracing the height of each outcome on a dot plot or bar graph (referred to earlier in Chapters 3 and 4 as the *over-fitted shape*), then any discussion based on comparing the *individual outcomes* of the model and sample distributions would be the same.
as comparing the shape, given the model and sample distributions are different, as they were in this task.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Description</th>
<th>Example excerpt</th>
<th>Number of teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual outcomes</td>
<td>The observed counts for each outcome in the sample data are similar/different to the expected counts for each outcome under the model</td>
<td>“…the model underestimates the number of hours where no emails are received (and somewhat for the number of hours where 4 emails have been received), and overestimates the number of hours where 1 email is received...” [Teacher 1]</td>
<td>8</td>
</tr>
<tr>
<td>Shape</td>
<td>The shape of the sample distribution is similar/different to the shape of the model distribution</td>
<td>“…this distribution is similar to the Poisson distribution in that: it is skewed to the right...” [Teacher 13]</td>
<td>7</td>
</tr>
<tr>
<td>Measures</td>
<td>Measures like the mean, median, mode, range and standard deviation are similar/different for the sample and model distributions</td>
<td>“…the mode of the Poisson distribution is 1, the mode for the emails is 0...” [Teacher 9]</td>
<td>4</td>
</tr>
<tr>
<td>Sampling variation</td>
<td>The differences between observed counts and the expected counts can (not) be explained by sampling variation, including sample size</td>
<td>“…the high frequency of zero and the low frequency of one are within what might be expected due to random variation in a sample from a Poisson distribution...” [Teacher 4]</td>
<td>3</td>
</tr>
<tr>
<td>Combined outcomes</td>
<td>The observed counts across a subset of outcomes in the sample data are similar/different to the expected counts across a subset of outcomes under the model</td>
<td>“…The fitted model has the number of emails received, 0, 1 and 2, as the most common which does fit that ‘skewed to the lower end’ aspect of a Poisson dist. as well...” [Teacher 12]</td>
<td>2</td>
</tr>
<tr>
<td>Alternative</td>
<td>The proposed model is (not) a better fit than an alternative model</td>
<td>“…a normal dist. bell shaded curve does not appear to match...” [Teacher 12]</td>
<td>1</td>
</tr>
</tbody>
</table>

However, teachers did not appear to recognise the connection between individual outcomes and shape in their descriptions, treating these features as unrelated. For example, Teacher 4 explained that s/he was “...concerned about the high frequency of zero and the low frequency of one...”, making use of the individual outcomes criterion, and then went on to explain “… but apart from those, the shape of the graph is what I would expect from a sample of 48...”, making use of the shape
criterion. This teacher appeared to disconnect the shape of the sample distribution from each outcome within the distribution and its associated frequency, by suggesting the shape of the sample distribution could be described without using two of the outcomes. By not recognising this relationship, Teacher 4 also did not appear to recognise that the observed counts for the outcomes of zero and one, which s/he is concerned about, could be explained by sampling variation.

Although teachers were instructed to use the visual fit of the model to the sample data to discuss the appropriateness of the proposed model, two teachers also discussed whether the proposed model would be useful and whether the conditions for the model were met as part of their reasoning. These considerations both link to the idea of applying the model beyond a data-fitting exercise.

5.6.4 Summary

The probability distribution model was central to teachers’ reasoning when integrating knowledge about the context, the probability distribution model and the sample data when describing their probability distribution modelling process. Teachers used a variety of criteria to sketch the shape of a distribution, and these criteria show that sample size is not taken into account and that other factors not seen in the data distribution are also used when shape sketching. Teachers also used a variety of criteria to assess the fit of a probability distribution model, most commonly comparisons of shape and of expected counts versus observed counts. However, these criteria appear unreliable as they did not allow teachers to correctly assess the fit of a probability distribution model.
5.7 Results from Task F: Using a simulation-based tool for probability distribution modelling

For Task F, teachers were guided through the use of the new simulation-based tool for probability distribution modelling and then asked to reflect on what they had learned. Nine of the 17 teachers in the study completed this task. These reflections highlighted three potential benefits - using the tool to reinforce a modelling perspective, to develop specific understandings for probability distributions, and to informally test the fit of a probability distribution model.

Using the tool to reinforce a modelling perspective. Teachers commented on how using the tool helped support the separation and connection between the real world and the modelling world, making statements such as:

... I really like how you can compare your experimental data to the modelled data side by side... [Teacher 5]

... The connection between the visuals (real life, simulated data from a theoretical distribution) are very clear ...... by being able to take the tracked shape and transfer it over to the real life data makes it easy to see the similarities and differences, as well as understanding where the tracked shape came from ... [Teacher 1]

Teachers were also positive about how the tool kept the focus on probability distributions as models and reinforced the notion that models were an approximation of reality:

... allows students to explore several different models easily ... [Teacher 3]

... the fitted model isn’t necessarily a perfect fit to the real data which is something students can struggle with ... [Teacher 12]

... the possibility of more than one probability distribution fitting the data can be seen ... [Teacher 9]

... the inexact nature of fitting a model ... [Teacher 3]
The design of the tool in terms of layout and interface was also seen by teachers as intuitive and the use of visualisation was perceived as being beneficial to student learning and engagement:

... it will be useful for students to simulate samples from a variety of distributions quickly ... seeing sampling variability with the shadowing allowing them to hold multiple iterations of a simulation in their head at once ... [Teacher 4]

... I have found that with other visualisation tools (e.g. the ones from iNZight) these help to engage students ... [Teacher 12]

**Using the tool to develop specific understandings for probability distributions.**

One barrier identified by teachers when using simulation to teach probability distribution modelling was limitations of available software, and the impact of these limitations on using simulation to build important ideas for probability distribution modelling. As discussed earlier in this chapter, teachers in the study generally viewed simulation as an important tool to visualise sampling variation, therefore the teachers were positive about how the tool could allow students to learn more about the features of different probability distributions through seeing the data generated from the models.

The comments made by teachers regarding specific understandings for probability distributions that use of this tool could support are summarised below:

- Visualisation of different probability distributions
- Impact of change of parameter(s) on the shape of the probability distribution
- Visualisation of randomness (variation within the distribution) through use of simulations
- Expectations for amount of distributional shape variation and the effect of sample size.
Using the tool to informally test the fit of a probability distribution model. Across the responses, teachers were in agreement that the tool could allow students to make more secure conclusions regarding the fit of a probability distribution model to sample data. This potential benefit was demonstrated by the responses given by Teacher 13. Presented with Version 2 of Task E (see Figure 29), this teacher had initially described that the proposed Poisson model was a reasonable fit to the sample data.

The Poisson model gives a reasonable approximation for the true distribution of emails ... [Teacher 13]

![Figure 29: Version 2 of Task E, which shows the sample distribution (n = 480) with the model probability distribution overlaid on the same graph.](image)

However, after using the simulation-based modelling tool in Task F (see Figure 30), this teacher was able to incorporate and visualise the variation associated with a sample distribution of size 480 when considering the fit of the model, as shown in the excerpt below:

Gives me a much, much better idea about the amount of expected variation between the experimental data and the model - much less than I expected! [Teacher 13]
Figure 30: An example of how the new simulation-based modelling tool could have been used by Teacher 13 to test the fit of the probability distribution model (see section 4.5.4 for more information)

Teacher 13 was not alone in reflecting on how the tool had helped enhance his/her personal understanding of probability distribution modelling, as seen in the excerpts that follow:

It has also really clarified for me that your data doesn’t have to fit the model exactly, as each simulation of the model will give slightly different results. As long as your data is within the range of values from running the simulation, then it is ok. [Teacher 5]

The amount of data collected impacts on the accuracy of the model selection. This is not something I recall explicitly teaching to date, and now I will. [Teacher 9]

5.7.1 Summary

Teachers were positive about potential benefits of the new simulation-based modelling tool, and identified ways that the tool could reinforce a modelling perspective and support development of understanding for features of probability distributions. Teachers also communicated that the tool had allowed them to take into account sample size when testing the fit of the probability distribution model through the visualisation of expected distributional shape variation.
5.8 Summary of results

Although teachers demonstrated that they understood the differences between the *distribution of true probabilities*, the *distribution of model estimates of probabilities* and the *distribution of experimental estimates of probabilities*, their knowledge for theory-driven approaches to probability distribution modelling appeared to be stronger than their knowledge for data-driven approaches. The use of experimental or sample data within the probability distribution modelling approaches demonstrated by teachers appeared problematic, specifically concerning methods such as sketching the shape of a sample distribution and testing the fit of a probability distribution model. In both of these methods, the amount of data available was either not used or not used consistently by the teachers. Before being introduced to the new simulation-based modelling tool, teachers did not clearly articulate how the use of simulation within probability distribution modelling learning contexts could reinforce the view of the probability distribution as a model not the reality. Teachers made positive comments about the new simulation-based modelling tool and reflections made suggest that use of the tool could allow for a more reliable *informal test of the fit of a probability distribution model*. 
Chapter 6
Discussion

The purpose of this study was to explore how teachers informally test the fit of a probability distribution. The main research question and supporting research questions are now answered in Section 6.1.

6.1 How do teachers informally test the fit of a probability distribution model?

Teachers used a variety of criteria to informally test the fit of a probability distribution model, most commonly comparisons of shape and of expected counts versus observed counts. There was high level of alignment between the criteria used by teachers for informally testing the fit of a probability distribution model and the criteria used in the example NZQA assessment item schedule (see Figure 1b). However, these criteria appear unreliable as they did not allow teachers to correctly assess the fit of a probability distribution model. The main issue appeared to be that in informally testing the fit of a probability distribution model, the size of the sample was not taken into account.

It should be noted that teachers in this study were following a method demonstrated in government produced resources, and the only guideline found by the researcher regarding sample size was that samples of at least 200 should show a shape similar to the underlying probability distribution. Despite this guideline, three out of five teachers made the incorrect call about the fit of the probability distribution model for Version 2 of Task E, which had a sample size of 480. It is surprising that teachers of statistics at curriculum level eight (Year 13) would not access such a key statistical inference concept as sample size when making their own “calls” about the visual fit of the proposed probability distribution to the sample data.

While curriculum achievement objectives should drive the teaching of statistics at curriculum level eight (Year 13), in the researcher’s experience what is taught is often limited to what is assessed by each achievement standard. The 2015 NZQA
achievement data\textsuperscript{10}, shows that the number of students assessed against AS91584 is just over a half of the number of students assessed against AS91586 (see Table 14).

<table>
<thead>
<tr>
<th>Standard number and title</th>
<th>Statistical inference methods assessed</th>
<th>Number of students assessed (2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS91582 Use statistical methods to make a formal inference</td>
<td>Confidence intervals for difference of two means/medians</td>
<td>13646</td>
</tr>
<tr>
<td>AS91583 Conduct an experiment to investigate a situation using experimental design principles</td>
<td>Randomisation test for difference of two means/medians</td>
<td>5128</td>
</tr>
<tr>
<td>AS91584 Evaluate statistically based reports</td>
<td>Confidence intervals for single and difference of two proportions</td>
<td>4791</td>
</tr>
<tr>
<td>AS91586 Apply probability distributions in solving problems</td>
<td>Informal test for the fit of a probability distribution model</td>
<td>9334</td>
</tr>
</tbody>
</table>

The significance of comparing the numbers of students in these two achievement standards is that AS91584 is the only achievement standard that requires students to construct and use confidence intervals for proportions. AS91584 covers statistical inference concepts with categorical data and probability models.

6.1.1 How are statistical inference concepts used when teaching probability distribution modelling?

Teachers in the study used a comparison of model counts (expected counts) and sample counts for outcomes (observed counts) as a criterion for testing the fit of the probability distribution model. This criterion is similar to findings from Lehrer, Jones, and Kim (2014), who found that students compared model statistics

with sample statistics as their criterion for model fit. However, comparisons of expected counts and observed counts need to take into account sample size.

Additionally, in order to use sample size as part of an informal test for the fit of a probability distribution model, teachers need knowledge of expected variation of either the model proportions or the sample proportions. It appears in this research that teachers had difficulty visualising this variation, similar to how the students in Dolor and Noll’s (2015) study lacked experience with expected variation for group proportions. The kind of thinking teachers in the study needed to use can be demonstrated through the animations of sampling variation developed by Wild et al. (2011)\textsuperscript{11}. In Figure 31, the two versions of Task E from the model are shown, with screenshots of visualisations of sampling variability for similar size samples for proportions of different modes of transport for students in the Census at School NZ databases.

![Two versions of Task E](image1)

**Two versions of Task E**

- \( n = 48 \)
- \( n = 480 \)

![Screenshots of visualisations of sampling variation](image2)

**Screenshots of visualisations of sampling variation**

Although the sample sizes used for the screenshots are not the same as the sample sizes used for Task E, they demonstrate roughly the amount of expected variation

---

\textsuperscript{11} See [https://www.stat.auckland.ac.nz/~wild/WPRH/index.html](https://www.stat.auckland.ac.nz/~wild/WPRH/index.html)
that, if visualised by teachers who completed Task E, could have helped teachers to make the correct call about the fit of the probability distribution model. The screenshots of the visualisations of sampling variation also show how the expected variation for proportions also depends on the size of the proportion as well as sample size.

The limitation of comparing the probability distribution model counts (expected counts) with the sample data counts (observed counts) extends to the comparison of distributional shapes. The use of shape as a criterion for probability model testing appears to be problematic. Firstly, the concept of distributional shape is not useful for probability distribution models with nominal categorical variables (for example the modes of transport data shown in Figure 31) and also for probability distribution models that are shaped irregularly. Secondly, from a probability modelling perspective, the researcher proposes that “shape” is a probabilistic interpretation of what outcomes are possible (using sample space) and which are more likely to happen or not (using proportions). When considering the criteria used for informal inferential reasoning involving numerical variables, shape is not used as a criterion for “making the call” (Wild et al., 2011). The inconsistency of the distributional shape sketches for small sample sizes demonstrated by teachers in the study perhaps indicates that a probabilistic interpretation of shape is challenging. Therefore, the researcher proposes that distributional shape is better explored by tracking the variation of proportions, as demonstrated in her new simulation-based modelling tool.

The use of shape also appeared problematic due to the different views teachers had on what the distributional shape sketch represented. Some teachers in the study viewed shape as ‘descriptive’ (the shape of the sample distribution they can see) while others viewed shape as ‘inferential’ (the shape of the underlying probability distribution that is not seen). The view that the distributional shape was inferential led to some teachers believing points could be added or removed from the sample data to get a better measure of the shape. This view was consistent with Casey and Wasserman (2015) who also found that teachers ignored points when fitting a line of best fit to bivariate sample data because they
used contextual knowledge about the population relationship. In a test for significance, although contextual knowledge should be used to interpret the results of the test, it is questionable whether contextual knowledge should be used as part of any criteria for the test.

6.1.2 How do teachers understand “distribution of experimental estimates of probabilities”?

When asked to describe an activity that could help students understand distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities, teachers gave correct explanations and provided good examples of teaching activities that would support the understanding of these terms. Consistent with government produced resources, and other researchers (e.g. Konold & Kazak, 2008; Nilsson, 2014; Roback et al., 2006), teachers discussed both simulated data and sample data as experimental or empirical distributions. However, when teachers described using simulations, it was not clear that they conceived a probability distribution as a model, instead using language associated with sample-to-population inference. Given that it has been found that when using simulation-based inference, students and teachers can confuse real sample distributions with simulated distributions (Gould, Davis, Patel, & Esfandiari, 2010; Pfannkuch, Wild, & Regan, 2014), it appears that greater care may need to be taken when using simulations in the teaching of probability distribution modelling. In the proposed statistical modelling framework developed by the researcher, simulated data sit in the model world and sample data sit in the real world and the word experimental is not used. Using a modelling framework such as this one could reinforce the notion that simulated data are generated from a model and real data are collected from the true situation being modelled. This appears to be an important distinction for teachers and students to understand, given that many probability modelling activities used by researchers use both simulated data and real data (e.g. Konold & Kazak, 2008; Lehrer et al., 2015).

6.1.3 How is the informal test of the fit of a probability distribution model taught within the wider probability distribution modelling process?
Teachers are expected to teach students to consider whether a probability distribution model is “good” based on both contextual knowledge and an informal test for the fit of a probability distribution model (see NZQA assessment item schedule, shown in Figure 1b). Contextual knowledge is used to check the conditions of the theoretical probability distribution model. Teachers are therefore required to combine a data-driven modelling approach (the use of data to test a model) and a theory-driven modelling approach (the use of the theoretical features of the probability distribution modelling). Pfannkuch and Ziedins (2014) proposed that teaching approaches to probability modelling that combined data-driven modelling and theory-driven modelling may obscure the purpose of modelling if the tasks used are not well-developed. The examples of tasks that Pfannkuch and Ziedins referred to were ones where the probability distribution models can be created theoretically and by using data (e.g. Fielding-Wells & Makar, 2015). Typically teachers use data, the experimental probabilities, as a vehicle to understand the theoretical probability distribution.

In this study, teachers’ considerations about the key understandings required for the learning AS91586 were mainly theory-driven. When describing their probability distribution modelling process, the probability distribution model was central to their reasoning when integrating knowledge about the context, the probability distribution model and the sample data. If this possible attribute of teacher thinking is combined with previous discussion regarding potential issues with viewing all data as being the same, then it may be that the product of a combination of data-driven modelling and theory-driven modelling approaches is the bi-directional reasoning discussed by Nilsson (2014). From the bi-directional perspective, the two main components for probability distribution modelling are the model and the data. Hence the distribution of true probabilities could be placed on the periphery, viewed as contextual knowledge that is integrated only at the model application stage, for example, when considering if the model is useful.

The researcher proposes that one of the reasons why combining data-driven modelling and theory-driven modelling approaches may result in collapse of two-world thinking to bi-directional thinking is that teachers need support to combine
statistical inference thinking with probability distribution modelling thinking. Figure 32 shows how the proposed statistical modelling framework could help promote clarity with using data-driven modelling and theory-driven modelling approaches, that keeps to the forefront the notion that a model is an approximation to reality (Pfannkuch & Ziedins, 2014).

![Statistical modelling diagram]

**Figure 32: Combining model-driven and data-driven approaches in the proposed statistical modelling framework**

### 6.1.4 A new informal test for the fit of a probability distribution model

The results suggest that the current method for informally testing the fit of a probability distribution model is not reliable. The difficulty with visualising expected variation of distributional shape (outcome proportions) appears to be the main issue. In pursuing an informal test for the fit of a probability distribution model, graphical inference methods where graphs are used as test statistics appears to have some basis in literature (see Hofmann, Follett, Majumder & Cook, 2012). Alternatively, a new informal test for the fit of a probability distribution model that allows students to visualise noise for individual outcomes in a distribution could build on previous research work in judging the fit of a model.
using shape (e.g. Fielding-Wells & Makar, 2015; Kazak & Konold, 2008; Lehrer et al., 2015).

Teachers made positive comments about the new simulation-based modelling tool and reflections made suggest that use of the tool could allow for a more reliable informal test of the fit of a probability distribution model. Teachers also communicated that the tool had allowed them to take into account sample size when testing the fit of the probability distribution model through the visualisation of expected distributional shape variation. The new informal test of the fit of a probability distribution model displayed all of the characteristics developed by the researcher for such a test (see 3.2.1), characteristics which were guided by informal inferential reasoning research (e.g. Garfield et al., 2008; Pfannkuch, 2006). Specifically, the new informal test of the fit of a probability distribution model:

- does not use formal procedures, methods or language e.g., test statistic, null hypothesis, chi-square, P-value;
- draws a conclusion about the goodness of fit of a probability distribution model by looking, comparing and reasoning from distributions of data;
- builds conceptual understanding of the goodness of fit of a probability distribution model; and
- provides foundations to make the procedures associated with the chi-square goodness-of-fit test more accessible.

6.2 Limitations of this research

As this study involved a small self-selected sample of 17 teachers, the results cannot be generalised to all teachers of AS91586. However, the results can be used to indicate potential issues and suggest areas which could need further research regarding model testing. It should be noted that four of the ten teachers who completed all tasks had statistics majors or equivalent, which is higher than would be typically found among Year 13 Statistics teachers. There are limitations with drawing conclusions about teachers’ knowledge of data-driven probability distribution modelling as much of what is currently assessed by AS91586 is theory-driven probability distribution modelling in the form of application
problems. The research does not provide a measure of the full teaching knowledge of teachers for probability distributions, which is wider than the tasks used in this study.

The data were collected for this study using an online structured self-interview and the researcher was unable to follow up unclear responses with teachers. Tasks were completed by each teacher in a range of environments. There is no way for the researcher to determine if they were completed independently, and so other unknown factors may contribute to what was described in responses. There may be other ways to test the fit of a probability distribution model without overlaying the model line on the data. Therefore this research only shows one specific method that could be used. Although a randomised experiment was used to test the effect of sample size on teachers’ conclusion for the visual fit of the probability distribution model, the group sizes were small (five each), and therefore any conclusions are tentative.

6.3 Recommendations for future research

There is a limited amount of research regarding tests for significance that use non-numerical test statistics. Hofmann, Follett, Majumder and Cook (2012) have explored the use of graphical inference across a large range of types of data and graphical representations using “crowdsourcing” but the researcher is not aware of other research in this area. Within the New Zealand context, apart from curriculum level five informal inferential reasoning (making a call based on whether one median is outside the “box” of the other, which could be considered a graphical inference), all other statistical inference methods, informal or formal, involve a numerical test statistic or the use of confidence intervals. More research is needed in the area of graphical inference to investigate if this method could help build more robust ideas about testing the fit of probability distribution models. The theoretical basis of the new informal test for the fit of a probability distribution model also needs further research. Theoretical aspects of the test need further evaluation and development, including but not limited to, simultaneous inference and robustness.
This research included the development of a new informal test for the fit of a probability distribution model and a simulation-based modelling tool needed to perform this test. Further research is needed regarding how to teach probability distribution modelling using the new test and whether use of the new test improves teachers and students’ understanding of probability distribution modelling. The new informal test for the fit of a probability distribution model is also an informal test for significance. Care needs to be taken when interpreting the test output, particularly as students can incorrectly interpret large P-values as evidence the null hypothesis (the probability distribution being modelled) is true (e.g. Reaburn, 2014). The new simulation-based modelling tool allows students to quickly change the parameters of a probability distribution model, and the tracking of over-fitted shapes for the simulated distributions would show more than one probability distribution would “fit” the sample data. Research could investigate whether learning that multiple models could “fit” the sample data helps students understand why they cannot accept the null hypothesis.

There is also a need for further research regarding the positioning of the new informal test for the fit of a probability distribution model within the New Zealand Curriculum for Statistics. Possible considerations include: the type of learning that leads towards the use of this test at curriculum level eight (Year 13); how to integrate the test within teaching probability distribution modelling; how the proposed statistical modelling framework assists the development of understanding; and how to align the test with the statistical inferential reasoning pathway.

Further research regarding sketching of distributional shape is also needed. One area that this research could explore is the use of software, such as the sketching tool developed and used in Task C and Task E of this study. The tool allowed for teachers to use their finger on touchscreen or a mouse to sketch the distributional shape. It might be possible with future research to establish a theoretical basis for the thickness of the line used to sketch the distributional shape in relation to the sample size (see Figure 33).
Another use of software to assist sketching distribution shape is to “crowdsource” sketches. This could involve a tool which allows students to sketch the shape of a distribution and then see how their sketch compares to others. Similar interactive online drawing tools have been developed\textsuperscript{12,13} in other areas and these ideas could inform research for sketching distributional shape.

### 6.4 Teaching and assessment implications

The simulation-based modelling tool was redesigned after completion of the data collection phase of the study, following feedback from the teachers who took part in the study and a statistician. Several minor changes were made, mostly concerning the use of colour to improve the effectiveness of the visualisations and animations. For example, the tool now uses two different colours to reinforce the two worlds: one colour for the real world/data and another colour for the model world/data. The tool is now hosted on the researcher’s website\textsuperscript{14} and freely available for teachers and students to use.

At the end of 2016, this research and the new simulation-based modelling tool were shared with teachers at the Auckland Mathematical Associations Statistics Teachers’ Day and the Otago Mathematical Association Teachers’ Day. Activities and notes created by the researcher to support use of the tool were made available

\begin{itemize}
  \item \textsuperscript{12} https://www.nytimes.com/interactive/2017/01/15/us/politics/you-draw-obama-legacy.html
  \item \textsuperscript{13} https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html
  \item \textsuperscript{14} http://learning.statistics-is-aweseome.org/modelling-tool/
\end{itemize}
on the Census at School (NZ) website\textsuperscript{15} and through a shared Google Drive\textsuperscript{16}. The researcher plans to present this research at the New Zealand Association of Mathematics Teachers bi-annual conference in October 2017.

The researcher also intends to develop and share additional teaching activities and notes to support teaching of probability distribution modelling using the tool. For example, the researcher has also begun to develop a virtual world where data can be collected for modelling (see \textit{Stickland modelling}\textsuperscript{17}). Another example is work begun with an undergraduate statistics student to develop a teaching task that uses live streaming video of real processes associated with famous landmarks.

One of the findings from this study was that teachers did not connect contextual knowledge about the situation being modelled with data collected from the situation. Therefore, it is recommended that students need to experience more real random processes to help written descriptions of the context “come to life”.

Pfannkuch and Budgett’s (2016) proposed strategies to increase student understanding of probability (p. 64) could be used to help teachers develop learning tasks, which are:

1. Incentivise students to engage in understanding the ideas;
2. Use visual imagery;
3. Allow students to play around with chance-generating mechanisms;
4. Develop strategies to enable students to link across representations including extracting information from word problems;
5. Use contexts that students can relate to.

Although the use of real data is not directly stated in this list of strategies, the researcher believes that using live video of real processes would combine points 1, 2 and 5 effectively.

\textsuperscript{15} \url{http://new.censusatschool.org.nz/resource/all-models-are-wrong-but-some-are-more-wrong-than-others-informally-assessing-the-fit-of-probability-distribution-models-as91586/}
\textsuperscript{16} \url{https://drive.google.com/drive/folders/0BxSVTvbi4l6TmNjUWRBbUs1Wjg}
\textsuperscript{17} \url{http://learning.statistics-is-awesome.org/stickland-modelling/}
The statistical modelling framework was designed with the intention of the framework being used directly with students. For example, when investigating a modelling situation, students could use the framework to organise their thinking by writing about each of the components and the connections between the components. The proposed statistical modelling framework has the potential to be used across curriculum level eight of the New Zealand Curriculum for Statistics. In particular, AS91585 Apply probability concepts in solving problems which refers to the three interconnected ways to think about probability. Figure 34 provides an example of using the statistical modelling framework to think about a problem from the NZQA AS91585 2015 exam.

Regardless of whether a distributional shape sketch is descriptive or inferential in nature, if a sketch is not the over-fitted shape, the research suggests that sketching through the data could improve ideas such as signal and noise, and of sampling variation around a true value (for each proportion). When sketching from left to right, the questions students could pose to themselves are “How much data do I have?” and “Do I really think that this outcome is more/less likely than the ones around it?”. Both of these questions would require experience with seeing variability of proportions/counts. Similar to the strategy used by Fielding-Wells
and Makar (2015), students could create dot plots on the same scales to represent a sample distribution from a population and then sketch the over-fitted distributional shape using overhead transparency sheets. These overhead transparency sheets could be layered on top of each other to show the variation in distributional over-fitted shape from sample to sample.

A final consideration is the assessment of an informal test for the fit of a probability distribution model. As AS91586 is currently externally assessed through a one-hour pen-and-paper exam, discussion is needed regarding how the test could be carried out under these conditions. A screenshot of the tool output could be included in the exam question. As an example, the corresponding test output for the assessment item presented in Figure 1 has been included in Figure 35.

Figure 35: NZQA AS91586 2015 Q2(b) model fit test using new simulation-based modelling tool

The use of a screenshot of the test output would remove the use of skills currently assessed, skills such as calculating the sample mean and calculating probabilities
from a probability model. Alternative questions would need to be asked to assess the level of statistical thinking, for example, *What if a different value for lambda was used in the model, could there be another Poisson model that fits the data well?*

### 6.5 Conclusion

This study was initiated following the identification of a problem with the current method for *informally testing the fit of a probability distribution model*, as assessed by AS91586. The main issues with the current method are the difficulty of taking into account (1) sample size and (2) the expected variation of sample proportions when comparing model proportions with sample proportions, without visual representations of the expected variation or without using a test for significance that uses a numerical test statistic and associated procedures. The researcher also found that teachers struggled to separate the real world from the model world in their thinking about probability distribution modelling, appearing to conceive of two components: model and data. To resolve the issue with the current method for *informally testing the fit of a probability distribution model*, the researcher used her proposed statistical modelling framework as the basis for the design of a new simulation-based modelling tool that allows students to perform *graphical inference*, resulting in a new *informal test for the fit of a probability distribution model*.
References


Gould, R., Davis, G., Patel, R., & Esfandiari, M. (2010). Enhancing conceptual understanding with data driven labs. In C. Reading (Ed.), *Data and context in


Appendix A

Tasks

Task A: Teacher background

<table>
<thead>
<tr>
<th>Q1</th>
<th>Please select your gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>Female</td>
</tr>
<tr>
<td>☐</td>
<td>Male</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q2</th>
<th>How many years have you been teaching?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Include this year if you are currently teaching</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q3</th>
<th>How many years have you taught AS91586 Apply probability distributions in solving problems?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Include this year if you have already or are currently teaching AS91586</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q4</th>
<th>Did you complete a major in statistics or operations research (or equivalent) as part of your tertiary study?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>Yes</td>
</tr>
<tr>
<td>☐</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q5</th>
<th>What type of professional development would benefit you the most in your teaching of statistics (statistical investigations, statistical literacy, probability)?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Select as many as apply from the list</td>
</tr>
<tr>
<td>☐</td>
<td>School-based sessions</td>
</tr>
<tr>
<td>☐</td>
<td>Workshops offered by local Mathematics association</td>
</tr>
<tr>
<td>☐</td>
<td>Personal reading or self-directed study</td>
</tr>
<tr>
<td>☐</td>
<td>University course (lecture based)</td>
</tr>
<tr>
<td>☐</td>
<td>Online course</td>
</tr>
<tr>
<td>☐</td>
<td>Other (please specify in Q7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q6</th>
<th>In your opinion, who would be best to lead professional development?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Select one from the list</td>
</tr>
<tr>
<td>☐</td>
<td>Another teacher at my school or another school</td>
</tr>
<tr>
<td>☐</td>
<td>Regional advisor</td>
</tr>
<tr>
<td>☐</td>
<td>NZQA</td>
</tr>
<tr>
<td>☐</td>
<td>Outside expert</td>
</tr>
<tr>
<td>☐</td>
<td>Other (please specify in Q7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q7</th>
<th>If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.</th>
</tr>
</thead>
</table>


Task B: Preparing to teach AS91586 Apply probability distributions in solving problems

Please do not consult other resources when completing this task.

Q1
What resources do you refer to as part of your preparation to teach a unit (or sequence of lessons) towards AS91586 Apply probability distributions in solving problems?
Select as many as apply from the list
- School-developed teaching materials such as a unit outline and associated teaching materials (e.g. shared powerpoints)
- NZQA material
- Personal reading or self-directed study
- Senior Secondary Guide material
- Cemus at School material
- Textbooks
- Other (please specify in Q6)

Q2
What do you consider are the three key concepts/ideas/understandings for this unit?
Please write about one concept/idea/understanding per box.

Q3
What are two aspects you have found students struggle with when learning this unit?
Please write about one aspect per box.

Q4
Describe one activity that you use in your teaching unit for AS91586 Apply probability distributions in solving problems that builds understanding of the distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities.
Q5
Explain why this activity builds understanding of the distribution of true probabilities versus distribution of model estimates of probabilities versus distribution of experimental estimates of probabilities.

Q6
If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.
Task C: Sketching shapes of distributions - Q1

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

![Graph](image1)

Task C: Sketching shapes of distributions - Q2

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

![Graph](image2)

Task C: Sketching shapes of distributions - Q3

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

![Graph](image3)

Task C: Sketching shapes of distributions - Q4

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

![Graph](image4)
Task C. Sketching shapes of distributions - Q5

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

Task C. Sketching shapes of distributions - Q6

Please sketch the shape of the distribution shown to you below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and login again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like. When you are happy with the shape you have drawn, press the Completed button.

Task C. Sketching shapes of distributions - Q7

Q7

If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.
**Task D: Assessing student understanding of probability distribution models**

The lengths of jelly beans produced by a confectionery company are currently modelled using a normal distribution, with mean 18.2 mm and standard deviation 0.58 mm. This model is used as part of the company's quality control procedures.

A random sample of 50 jelly beans was taken from the production line and the lengths of these jelly beans are shown in the graph below.

![Graph showing lengths of jelly beans]

Two students are asked to use this data to consider whether the normal distribution model defined above is a good model for the lengths of jelly beans produced by this confectionery company.

- **Student A** sketches the shape of the sample distribution of jelly beans lengths (negatively skewed) and compares this to the shape of the model normal distribution (symmetric). As the shapes are not similar, the student concludes that the normal distribution is not a good model for the lengths of jelly beans produced by the confectionery company.

- **Student B** uses the sample data to calculate the proportion of jelly beans in the sample that are less than 19 mm in length. The student then uses the model defined above to calculate the proportion of jelly beans that could be expected to be less than 19 mm in length. As the proportions are similar, the student concludes the normal distribution is a good model for the lengths of jelly beans produced by the confectionery company.

**Q1**

For each student, (a) comment on what misunderstandings each student demonstrates about applying probability distributions and (b) briefly outline what further explanations or learning activities you would use to develop correct understanding.

Please write about one student per box.

---

**Q2**

If not discussed already, describe what else you would want students to discuss when asked to consider whether a probability distribution is a good model for a situation/variable.

---

**Q3**

If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.

---

**Task E: Capturing your modelling process - Q1 to Q2**

The number of emails received each hour to an email account was recorded over a 48-hour period.

This data recorded is shown below:

```
4, 3, 1, 0, 0, 0, 2, 0, 1, 2, 3, 2, 1, 1, 7, 4, 0, 3, 4, 2, 4, 2, 1, 0, 0, 2, 0, 1, 0, 0, 6, 0, 1, 3, 3, 2, 3, 0, 2, 2, 1, 3, 4, 2, 1, 2
```

You have been asked to investigate whether the number of emails received per hour to this email account can be modelled by a Poisson distribution.

**Note:** You do not have to carry out the investigation.
Q1
Please describe the steps that you would work through to carry out this investigation (which includes reporting your findings). You should draw on any statistical knowledge you have for probability distributions.

Your descriptions for each of these steps should include reference to any expectations, calculations, tests, drawings, graphs, models or theoretical probability distributions, conditions, assumptions or limitations you would use or consider as part of your investigation.

Please write about one step per box. Use as many of the boxes as you want to show each of your steps. You do not need to use them all.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Step 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q2
If you were to use this investigation with students, what issues (if any) do you anticipate students would have?
Task 1: Capturing your modelling process - Q3 to Q5

Q3
Please sketch the shape of the distribution of number of emails received per hour on the graph below.

Draw directly on the graph using a mouse (if using a computer) or your finger/stylus (if using a device). You can end this session and log in again using another device with a touchscreen if you find sketching the shape with a mouse difficult. You can re-draw the shape as many times as you like.

Q4
What criteria did you use for sketching the shape you drew in question 3? Or, in other words, how did you decide how to sketch the shape?

Q5
In reference to the shape you drew in question 3, do you think that most of your students would draw a similar shape? Why or why not?

Task 2: Capturing your modelling process - Q6 to Q8

A Poisson model has been used to generate theoretical (expected) counts for the number of emails received per hour, using \( \lambda = 1.77 \) (the mean number of emails received per hour for the 48-hour period). This theoretical distribution has been displayed on the graph below as a line graph (in yellow).

Q6
Discuss the appropriateness of the Poisson model for the number of emails received to this email account in terms of the visual fit of the model to the sample data.
Q7
In reference to your answer to question 6, do you think that most of your students would discuss similar things that you did when considering the appropriateness of the model in terms of the visual fit of the model to the sample data? Why or why not?

Q8
If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.

Task F: Using a simulation-based tool for probability distribution modelling

For AS91586 Apply probability distributions in solving problems, students are expected to compare the features of a distribution of model (theoretical) estimates of probabilities with a distribution of experimental (data) estimates of probabilities in order to discuss the appropriateness of a probability distribution model.

One aspect of this informal model fitting process that may cause issues is how students take into account the amount of data they have to fit/test a model against. This task will guide you through exploring a simulation-based tool for assisting an informal model fitting process. You will be then asked to answer questions on your experience of using this tool.
Task F: Using a simulation-based tool for probability distribution modelling

Welcome to this guided exploration of a simulation-based tool for probability distribution modelling.
As we move through this exploration, text will appear in this box first, along with demonstrations of the tool in action.
You will not be able to interact with the tool until the end of the guided exploration.
Click the NEXT button to move forward.

Next  Skip to the end and explore the tool

End session

Task F: Using a simulation-based tool for probability distribution modelling

You should see the interface for the tool appear below this box.
On the left-hand side is where we will think about and look at real data from the situation we are investigating.
On the right-hand side is where we will define our model and look at simulated data from this model.
We will focus on how to use the tool rather than demonstrating a full process to use for modelling.

Next

Task F: Using a simulation-based tool for probability distribution modelling

Suppose that someone was worried that a six-sided die they had bought from the $2$ shop was a bit shonky. They have rolled it twenty times and it has come up as six on 9 of those times.

Next

Task F: Using a simulation-based tool for probability distribution modelling

For this situation, we have real data that we can use to help develop the model. Data can be added into the tool by typing in values separated by commas, or by copying and pasting values that are separated by commas, or by copying a column of data from a spreadsheet.

Next

Task F: Using a simulation-based tool for probability distribution modelling

When data is entered, as long as the variable has been described, a dot plot of the data will appear.

Next

Task F: Using a simulation-based tool for probability distribution modelling

We want to compare this to data from a fair die. So we need a model - in this case a uniform discrete random variable.

Next

Task F: Using a simulation-based tool for probability distribution modelling

We need to set the lowest possible value to 1 and the highest to 6. We'll set the number of trials to match the amount of real data - so 20, since we want to simulate 20 dice rolls from a fair die.

Next
Task F: Using a simulation-based tool for probability distribution modelling

Clicking the Run (another) simulation button will generate data randomly from this model and add it to the graph directly below.

The two graphs - real vs simulated data - are set to have the same properties, which can be adjusted at the bottom of the page. In this case, we’ll slide the size of the dots a bit smaller (this might not always be necessary). You can do other adjustments here - like change the min and max values of the scale, or round the values (helpful for continuous random variables).

Clicking the Start animation button will animate repeated simulations.

To capture the shape of the model that takes into account the amount of data we have, click the the option Track shape for simulated data below the graph of the simulated data.

Let the animation play for a while to build up a visual picture of the variation we could expect when using this model with this amount of data.

Clicking the Stop animation button will stop the animation and we can take a closer look at how the real data compares to the simulated data. There are various options to capture features of the data (you can try them later) but we’ll focus on the shape.

Clicking the the option Transfer over shape of model below the graph of the real data allows us to compare the model (along with variation) to our real data.
Task F: Using a simulation-based tool for probability distribution modelling

Go to feedback questions

A 'poor' fit for a model could be indicated if you can see the 'tops' of the 'stacks' of dots for the real data - either above or below the shape shaded area.

Let's try another exploration!

Next

Task F: Using a simulation-based tool for probability distribution modelling

Go to feedback questions

This next situation is based on the work by the British statistician R.D. Clarke (An Application of the Poisson Distribution, 1946).

During World War II, London was assaulted with German flying-bombs. The British were interested in whether or not the Germans could actually target their bomb hits or were limited to random hits with their flying-bombs. The British mapped off the central 24 km by 24 km region of London into 586 1/2 km by 1/2 km square areas. They recorded the number of bomb hits in each of the areas.

Next

Task F: Using a simulation-based tool for probability distribution modelling

Go to feedback questions

We could model this situation with a Poisson distribution, using the mean of the bomb data as an estimate for \( \lambda \). To get the mean of the bomb data, clicking the Show mean button under the graph.

Next

Task F: Using a simulation-based tool for probability distribution modelling

Go to feedback questions

When tracking the shape for the simulated data, there is not as much variation visible as with the fair dice model.

Next

Task F: Using a simulation-based tool for probability distribution modelling

Go to feedback questions

This time, the Poisson model looks like a good fit to the bomb hits data. Note, this does not mean that we can say the bombs were being randomly tossed, just that the real data looks similar to data generated from a Poisson model.

Next

You can explore the tool more here. To get you started, you should find the email data from Task E already copied into the tool.

Once you have explored this situation, you can reset the tool and try simulating data from different models or copying in any real data you have.

When you have finished exploring, click the Go to feedback questions button.
Task F: Using a simulation-based tool for probability distribution modelling

Q1
What benefits (if any) do you believe using this tool would have for building students' understanding of informally fitting a probability distribution model?

Q2
What key concepts/ideas/understandings do you believe this tool highlights for students?

Q3
What understandings about informally fitting a probability distribution model, if any, has using this tool helped clarify for you?

Q4
Please access the Senior Secondary Guide activity Lateness: Choice or Chance by [linking on this link] (this link will open in a new window). Review the activity and then return to this page. Have you read and used any ideas from this activity in your classroom before today? Please discuss.

Q5
If you would like to make any further comments about this task, please do so here. This could include clarifying any responses or further thoughts/comments related to the task questions.
Appendix B

Ethics related documents

Office of the Vice-Chancellor
Finance, Ethics and Compliance

UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE (UAHPEC)

07-Mar-2016

MEMORANDUM TO:
Assoc Prof Maxine Pfannkuch
Statistics

Re: Application for Ethics Approval (Our Ref. 016899): Approved with comment

The Committee considered your application for ethics approval for your project entitled Teachers’ knowledge of probability distribution modelling.

Ethics approval was given for a period of three years with the following comment(s):

Thank you for a well-written application.

1. Please ensure that the UAHPEC approval wording on the Advertisement is the correct form: Approved by the University of Auckland Human Participants Ethics Committee on _____________ for three years, Reference Number _____________.
2. On the CF, in the third bullet point, ‘understand’ should be ‘agree’.

The expiry date for this approval is 07-Mar-2019.

If the project changes significantly you are required to resubmit a new application to UAHPEC for further consideration.

In order that an up-to-date record can be maintained, you are requested to notify UAHPEC once your project is completed.

The Chair and the members of UAHPEC would be happy to discuss general matters relating to ethics approvals if you wish to do so. Contact should be made through the UAHPEC Ethics Administrators at ero-ethics@auckland.ac.nz in the first instance.

All communication with the UAHPEC regarding this application should include this reference number: 016899.
PARTICIPANT INFORMATION SHEET

To: Year 13 statistics teachers with experience teaching AS91586

Project title: Teachers' knowledge of probability distribution modelling

Name of researcher: Anna-Marie Martin

Supervisor: Maxine Pfannkuch

My name is Anna-Marie Martin and I am a Professional Teaching Fellow in the Department of Statistics at the University of Auckland. Currently I am undertaking a dissertation towards a Master of Professional Studies (Mathematics Education) at The University of Auckland. This requires me to conduct research related to Mathematics Education.

I would like to invite you to participate in my research, which will explore teachers’ knowledge of probability distribution modelling. In particular, my research will focus on teacher knowledge of the curriculum achievement objectives associated with AS91586 Apply probability distributions in solving problems.

The overall aim of this project is to capture good teacher practice and use this to create a framework of key concepts for the teaching of probability distribution modelling. Analysis of teacher responses to tasks will also help identify what areas, if any, teachers of probability distribution modelling need further support and guidance with to ensure student learning of probability distribution modelling processes promotes good statistical thinking. The findings of this project will be used to inform future research into the area of teaching probability modelling and to guide professional development of pre- and in-service statistics teachers.

I wish to collect data for the study through a sequence of six tasks that will be completed in an online environment. These tasks will ask you various questions related to teaching AS91586 Apply probability distributions in solving problems. You will also be asked questions related to your teaching experience and your opinions on professional development for statistics teachers. As part of completing these tasks, you will have access to a new online tool designed to help develop student understanding of probability distribution modelling. You will also be able to contribute to, and benefit from, the creation of a common body of knowledge for teaching probability distribution modelling. You may also benefit from reflecting on your own practice as a result of completing the tasks.

These tasks should take you no more than four hours to complete. You can complete these tasks in multiple sessions over a period of no less than two weeks, at whatever time and place suits you. These tasks will be made available to your through a secure website. To participate in this study, you will need to agree not to share your login details with any other person. You will also need to agree to complete the tasks independently and privately. I will also require your assurance that you will not make copies of or share the tasks you will complete in this study.

I will be aware that you are participating in this research because you have contacted me directly to express your interest, therefore your participation is not anonymous. I give my assurance that your decision to participate or to not participate in this research will not affect any relationship I may have with you. I guarantee to maintain a professional relationship with you. The login details you will use to access the secure site will be system generated.
and will not be known to anyone else apart from you. This means your completed online tasks are anonymous and I will not be able to associate your responses to the online tasks with you directly.

If you agree to participate, all the information I obtain will be reported or published in such a way that neither you nor your school will be identified. All data will be kept secure in locked or password protected files for six years and then destroyed, as is usual for research at this university. Paper files will be shredded and computer files deleted. The research findings will be used in my Master of Professional Studies dissertation, and may also be used in journal articles, conference presentations and for the development of professional development material including workshops. You may request a copy of the summary of the research findings by indicating on the attached consent form.

Participation in this research is completely optional. You may withdraw your participation any time without giving a reason. As the online tasks will be completed anonymously, to have your responses to the tasks deleted, you will need to contact the data manager, Marie Fitch mfitch@auckland.ac.nz, and supply your login details so she can delete your responses. You are free to withdraw any data related to you any time up to 30th May 2016 without giving a reason. The data manager will sign a confidentiality agreement not to disclose any information provided to a third party.

If you are willing to participate in this study, please fill in the consent form and return this to me via email. Following your consent, I will arrange for access to the secure site and send you instructions.

If you have any questions please feel free to contact me:

Anna-Marie Martin

Email: anna-marie.martin@auckland.ac.nz

My supervisor is: The Head of Department is:
Associate Professor Maxine Pfeiffer Associate Professor Ilse Zieler
The University of Auckland The University of Auckland
Department of Statistics Department of Statistics
Private Bag 92019 Private Bag 92019
Auckland 1142 Auckland 1142
Phone: +64 9 923 8794 Phone: +64 9 923 5051
Email: m.pfeiffer@university.ac.nz Email: i.zieler@university.ac.nz

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. 89711.

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE on 7th March 2016 for a period of three years. Reference Number 016899
CONSENT FORM

THIS FORM WILL BE HELD FOR A PERIOD OF SIX YEARS

Project title: Teachers' knowledge of probability distribution modelling
Name of researcher: Anna-Marie Martin
Contact email address for researcher: anna-marie.martin@auckland.ac.nz
Supervisor: Maxine Pfannkuch

I have read the Participant Information Sheet. 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 take part in this research.
- I understand that I will complete a sequence of six online tasks related to teacher knowledge of probability distribution modelling, with a total time commitment of around four hours over two weeks.
- I agree that I will not share my login details to the secure site that hosts the online tasks with any other person.
- I understand that I will complete the tasks independently and privately.
- I understand that I may not make copies of or share the tasks completed in this study.
- I understand that I am free to withdraw participation at any time without giving a reason.
- I understand that I am free to withdraw any data related to me at any time without giving a reason up to 30th May 2016.
- I understand that data will be kept for six years, after which time any data will be destroyed.
- I understand that neither I nor my school will be identified in any reports.
- I wish to receive a summary of findings, which can be emailed to me at this email address: 

Name: __________________________
Signature __________________________ Date __________________________

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE on 7th March 2016 for a period of three years. Reference Number 016899
### Appendix C

**Additional results**

*Table 15: Terms used when discussing each type of probability distribution, from a review of the documents written by Hinchliffe and Priest (2013) and Lakeland and Nugent (2016)*

<table>
<thead>
<tr>
<th>Distribution of true probabilities</th>
<th>Distribution of model estimates of probabilities</th>
<th>Distribution of experimental estimates of probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>True probabilities</td>
<td>Probabilities calculated using theoretical models</td>
<td>Experimental estimates</td>
</tr>
<tr>
<td>Usually unknown</td>
<td>Probability model for the population</td>
<td>Long-run proportions of occurrences in an experiment</td>
</tr>
<tr>
<td>Unique to situation</td>
<td>Theoretical probabilities</td>
<td>Experimental probability distribution</td>
</tr>
<tr>
<td>What would actually happen in real world</td>
<td>Model estimates of probabilities</td>
<td>Sample from a population</td>
</tr>
<tr>
<td>Population probability distribution</td>
<td>Theoretical distribution</td>
<td>Sample distribution from sample</td>
</tr>
<tr>
<td>Population distribution</td>
<td>Model</td>
<td>Probability distribution from sample</td>
</tr>
<tr>
<td>True distribution of probabilities</td>
<td>Theoretical model</td>
<td>Distribution of experimental probabilities</td>
</tr>
<tr>
<td>Actual situation</td>
<td>Probability model</td>
<td>Long-run relative frequencies</td>
</tr>
<tr>
<td>True data</td>
<td>Probabilities for the theoretical model</td>
<td>Simulation results</td>
</tr>
<tr>
<td>Real life distribution of data</td>
<td>Theoretical model</td>
<td>Simulation probabilities</td>
</tr>
<tr>
<td>Data usually unknown</td>
<td>Theoretical results</td>
<td>Experimental estimates of probabilities</td>
</tr>
<tr>
<td>Actual distribution</td>
<td></td>
<td>Sample data</td>
</tr>
<tr>
<td>Real life probability distribution</td>
<td></td>
<td>Simulation data</td>
</tr>
<tr>
<td>True probability distribution</td>
<td></td>
<td>Experimental data</td>
</tr>
<tr>
<td>Situation</td>
<td></td>
<td>Probability distribution</td>
</tr>
<tr>
<td>True data distribution</td>
<td></td>
<td>Data from the sample</td>
</tr>
<tr>
<td>True distribution</td>
<td></td>
<td>Data from the simulation</td>
</tr>
<tr>
<td>Real distribution</td>
<td></td>
<td>Experimental distribution</td>
</tr>
<tr>
<td>Distribution</td>
<td></td>
<td>Sample distribution</td>
</tr>
<tr>
<td>True situation</td>
<td></td>
<td>Experimental results</td>
</tr>
<tr>
<td>Real situation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

156