A vision for statistics and data science education at the senior secondary level in Aotearoa New Zealand: Statement from the New Zealand Statistical Association Education Committee, December 2021.

Our statement summarises our main perspectives on statistics and data science education at the senior secondary level (specifically Year 12/13), with a focus on considering the necessary learning to support a transition from high school to first-stage tertiary level statistics or data science courses. There are multiple perspectives that should be considered when defining "what matters the most" for statistics and data science education and we do not present this statement as being inclusive of all perspectives. For example, statistics and data science education needs to include Mātauranga Māori and Te Ao Māori approaches to data (see Kylie Reiri's recommendations[1]). We have drawn on our areas of expertise, informed by international trends for education, current research literature, university perspectives and a review of the teaching and learning promoted by the current curriculum and achievement standards. We hope to provide some answers to the following question: *What matters the most for statistics and data science education at the senior secondary level in Aotearoa New Zealand?*

1. Statistics and data science education - these are not the same thing for Aotearoa New Zealand

We refer to both statistics education and data science education in this statement. Internationally, there is no consensus on how statistics and data science education should interrelate at the school level. In countries where statistics is not a significant area of the national curriculum, a description for school-level "data science" can often look like the statistics that we (in Aotearoa) already teach (e.g., YouCubed Data Science). However, it is important to note that any thought that "statistics education is the same as data science education" is not consistent with the positions adopted by universities within Aotearoa New Zealand, where nearly all offer separate majors, specialisations and qualifications in data science, alongside those in mathematics and statistics. In particular, all include substantial computational components that extend far beyond anything covered in statistics, in line with international trends. Therefore, it will be important to identify helpful similarities and differences for statistics and data science education with respect to curriculum, pedagogy and assessment (see also Appendix B).

1.1 What's happening internationally in statistics and data science education research?

Statistics and data science education research continues to explore important areas of teaching and learning. The most comprehensive single source is the Springer book edited by Ben-Zvi., Makar & Garfield (2018). While concentrating on statistics education there is quite a lot of attention given to data science emphases and many of the authors are working across both fields. Data science education research and relationships with statistics education research can be seen in a number of very recent special issues for journals and the themes for international conferences, for example:

- Statistics Education Research Journal - Special Issue on Data Science Education (2022)
- Teaching Statistics - Special Issue on Teaching Data Science and Statistics (2021)
- Journal of Statistics and Data Science Education Special Issue: Computing in the Statistics and Data Science Curriculum (2021)
- International Association of Statistics Education - Satellite Conference on Statistics Education in the Era of Data Science (2021)

[1] https://www.cio.com/article/3629466/adopting-a-te-ao-maori-approach-to-data-and-analytics.html

A persistent finding for statistics education research is that hypothesis testing and confidence intervals are difficult for students and give rise to many misconceptions (e.g Case et al., 2018). Simulation-based inference appears to have pedagogical benefits as it adds to students conceptual development by giving them a stochastic conception of inference and an appreciation of the probabilistic nature of inference (e.g. Tintle et al., 2018). For example, bootstrapping supports the construction and understanding of interval estimation, and simple binomial experiments are commonly used to introduce reasoning with chance models via simulation approaches. With respect to developing statistical modelling, visual approaches are encouraged that support learners to develop concepts, aided by interactive and dynamic computational tools (e.g. Burrill, 2020). Research concerning modelling perspectives to probability (e.g. Kazak & Pratt, 2021) suggest theoretical and data-oriented perspectives of probability can be combined for teaching and could help connect probability to students' lives using modern and technological contexts.

In response to the "data deluge", what it means to be "statistically literate" is a current theme to research. Data is not just obtained from traditional study designs such as surveys or experiments, and models are not just those for inferring population parameters. Data literacy has emerged as a desirable goal for education (e.g. Gould, 2017), complemented by a need to focus more on civic responsibilities and social implications of data (e.g. Engel, 2019). With respect to data science education, new frameworks are being developed to guide teaching and learning (e.g. Fries et al., 2021), and suggest that a broader approach is needed than "one" statistical enquiry cycle (e.g. Cook et al., 2021), particularly with respect to modelling and ways of thinking (e.g. Gould, 2021). Emerging research highlights the need to carefully consider task and tool design when introducing new digital technologies or modelling approaches, such as coding and algorithmic models (e.g. Biehler & Fleischer, 2021). The need for significant specialist professional development for statistics and data science teachers remains crucial (e.g. Bargliotti & Franklin, 2015).

Appendices C and D provide some examples of research and collaborations undertaken by statistics and data science education researchers within Aotearoa New Zealand.

1.2 What's happening internationally for curriculum development?

To gain some international perspectives on statistics and data science at the senior secondary level, we have reviewed international high school statistics and data science curriculum projects. The diversity of modern interpretations for statistics and data science education at the senior secondary level can be seen in the emerging examples of international curriculum development. Some examples include: GAISE2, ProCivicStat, YouCubed Data Science, International Data Science in Schools Project, Bootstrap Data Science, IDC4U, Introduction to Data Science, CourseKata Statistics and Data Science, ProDaBi (Germany). There are also a number of initiatives that promote data literacy at the school level, including International Data Science in Schools Project, Data Science for Everyone and The Messy Data coalition, all of which are gaining support from a wide range of national statistics associations, international educators, businesses and local governments.

1.3 What's happening in first stage university courses in NZ?

We have also considered first-stage statistics and data science courses currently offered at universities within Aotearoa New Zealand. Unsurprisingly, there is also no consensus curriculum for these courses, however, discussions with a range of first-stage statistics and data science educators from across the universities suggests there are many similar features that could inform the senior secondary level. These include: exploratory data analysis; study design; formal inference with confidence intervals and hypothesis testing; simple linear regression; and probability methods and distributions. For courses that also teach data science, common features appear to be: the use of tasks that require students to engage directly with digital data sources such as APIs; students working closely with data using computational tools such as spreadsheets and programming languages (e.g., R); the consideration of data ethics, ownership, responsibility, and privacy; and student-led creation of a wide range of data visualisations and products.

2. A review of the teaching and learning promoted by the current curriculum and achievement standards

To answer the question, *What matters the most for statistics and data science education at the senior secondary level in Aotearoa New Zealand?,* we realised that it was important to consider and review the teaching and learning promoted by the current curriculum and achievement standards at the senior secondary level for statistics. To think about the future requires careful reflection on the past, in order to (1) avoid making the same missteps that may have hindered student learning, participation and achievement, and (2) to ensure teachers are supported to teach and assess the curriculum with confidence. As the curriculum element that always dominates all others is the assessed curriculum, we have gone carefully through the current Level 2 & 3 achievement standards (our assessed curriculum) and tried to address what is excellent about them and what can be improved.[2]

2.1 What's working well?

Aotearoa New Zealand's current statistics curriculum has been internationally recognised as world-leading, and is envied and emulated by many countries[3]. In fact, many features of international curriculum documents and first-stage statistics and data science courses are already being taught in our high schools. In large part, this is due to the long history of statistics education research undertaken within Aotearoa New Zealand (see Pfannkuch, M., Wild, C., Arnold, P., & Budgett, S., 2020). Rosemary Hipkins provided a summary of the progress made for statistics education within Aotearoa New Zealand in the TLRI report *Doing research that matters: A success story from statistics education*[4] and with Alex Neill a literature review of mathematics and statistics education in the 2015 [Review of Review and Maintenance Programme (RAMP) report](#) (Ministry of Education).

Some of the many positive aspects the teaching and learning promoted by the current curriculum and achievement standards include: students doing real investigations with real data; the use of multivariate data sets; students posing investigative questions from a data set; the use of open-book projects rather

[2] We should note, however, that there are important information sources that are not available to us, most notably what has been seen by NZQA moderators of elements that have caused serious problems.
[3] Sir David Spiegelhalter discussing statistics education in Aotearoa New Zealand on RNZ: How statistics can help us separate fact from fiction https://www.rnz.co.nz/audio/player?audio_id=2018712277refs (from 26mins 30secs)
[4] Rosemary Hipkins (2014). Doing research that matters: A success story from statistics education. http://www.tlri.org.nz/sites/default/files/projects/TLRI_Project_Plus_Web_0.pdf

than one-hour timed closed book assessments; clearly defined learning trajectories across multiple years for sample-to-population inference; wide-spread adoption of technology to facilitate learning from data; the specific teaching and assessment of statistical literacy; development of student understanding of inference using visual inference tools; and an interpretive focus on learning from data and telling the story of the data.

Broader factors that currently support effective teaching and learning include the careful consideration of when to combine different statistical ideas or methods. For example, currently confidence intervals are used for inferences involving sample situations and randomisations tests are used for inferences involving causation based on experiments. The approach of keeping these methods “separate” provides fewer complications in early experiences with inferential tools, and helps to underline the crucial differences between sample-to-population and experiment-to-causation generalisations.

2.2 What’s not working that well?

We must also consider what we can do better to ensure we remain forward-thinking and world-leading with our high school statistics education. Therefore, our review also identified some areas that need to be carefully considered with respect to future curriculum and assessment development.

The predominant use of written reports for assessing statistics
While reporting on investigations is an important part of statistics and data science, assessment opportunities are also needed for conceptual understanding and this has been largely absent. It may be that this partially resolved through the changed nature of the achievement standards as part of the Review of Achievement Standards (RAS) process, for example, that two of the four standards at each NCEA level within a subject will be externally assessed. Additionally, expectations for communication (written or otherwise) need to prioritise quality and clarity of statistical thinking over quantity.

Narrowly-defined internally-assessed achievement standards for statistics
Many of the internally-assessed achievement standards focus on a specific type of analysis or data-type rather than support selection from a set of appropriate statistical methods. This has led to severe constraints on the data that can be used, a paucity of opportunities for small curiosity-driven explorations, a small number of experiences with statistical graphics applied to data (preventing the learning that can only come from having many experiences), and a minimal focus on reasoning with categorical data beyond two-way tables and risk calculations or interpretations.

Impact of task structures and/or frameworks
A challenge for supporting the teaching and assessment of statistics and data science is to provide an appropriate structure/framework for a task that clarifies or reinforces key concepts but does not restrict creativity nor promote procedural approaches. Consequently, a balance is needed between assessment requirements that are too broad to support effective implementation, and those that are too narrow or specific and lead to “tick box” approaches. Additionally, there may not be one structure or framework that works for all tasks. Holistic approaches to assessing student understanding may require a different approach to assessment criteria and further teacher development.

Contextual knowledge should include personal and cultural knowledge or perspectives
There is a current expectation for “research” to be cited as part of reporting on statistical investigations and this has the potential for under-valuing of personal and cultural contextual knowledge or perspectives. The curriculum does not actually require “research” but contextual knowledge. The key idea for “contextual knowledge” is that students situate their explorations and investigations within a meaningful context, and that appropriate connections are made between the context, questions, data and its messages. We recommend extreme care with requirements around “using relevant” or “informed” contextual knowledge for future curriculum and assessment development, to allow for students to use and make personal connections with data (e.g. Jones, 2019).

Greater clarity is needed with respect to the learning progressions for probability
Attention is needed with respect to teaching mathematical probability alongside probability as a tool for statistical modelling, specifically inference, and how these dual purposes for probability develop across the different curriculum levels. Students do need to develop understanding of the mathematics of probability, including the use of mathematical notation and representations, and approaches such as expectation algebra. However, the teaching of probability also needs consideration of uncertainty, data, and use of probability models to provide estimates of true probabilities. Emerging research about using technology to support generalisations about probability structures or mathematics should inform future curriculum and assessment development.

2.3 What could be introduced or reimagined?

Our review of the current curriculum and achievement standards also identified many opportunities to strengthen the teaching and learning of statistics and data science at the senior secondary level in Aotearoa New Zealand. Here we have drawn on previous comments within this section, and our earlier considerations of how statistics and data science education is currently positioned within Aotearoa New Zealand.

Exploratory data analysis
Exploratory data analysis features in the statistics achievement objectives at both curriculum level seven and eight but is not a feature of current achievement standards. Exploratory data analysis involves curiosity-directed exploration of available data and typically involves looking at a large number of features and relationships in a short period of time in pursuit of unexpected insights and conjectures for follow-up investigation. As a consequence, the nature of the data explored can be broadened as analysis is not restricted to traditional analytical approaches. For example, students could explore the design of music album covers, define variables based on visual attributes, compare these over different decades or musical genres using a wide variety of visualisations, make conjectures about relationships, and develop interactive documents that communicate key findings from their exploration.

Greater use of technology for teaching probability
The use of technology for teaching probability concepts and modelling approaches could be expanded. We note that the annotated exemplars for AS91268 Investigate a situation involving elements of chance using a simulation[5] show the use of “by hand” methods involving calculator-based random number

[5] https://www.nzqa.govt.nz/ncea/subjects/mathematics/exemplars/level-2-as91268/

generators. Probability simulations at the senior secondary levels should involve thousands of trials and the careful design and use of computational tools is crucial to facilitating much bigger scale experiences with simulation. There is also scope to build on the initial work of the NZQA digital assessment project[6]. Furthermore, the use of explorations of “what if” scenarios for chance situations, through technology, could be used to develop probability understanding.

Broadening notions and awareness of data

The current curriculum achievement objectives refer to surveys, experiments and existing data sets. While these descriptions can be creatively interpreted to include a wide range of modern data sources, it would be better to specify these. Creating data from images, sounds, and movements, using web scraping and APIs, querying databases, generating data through interactions with digital devices, and using social media data, often involves greater computational skills than currently taught. The analysis of and interpretations from the data obtained also may require new understandings that don’t currently exist within our traditional observational vs experimental, or primary vs secondary data, perspectives. Additionally, data can be structured in many ways and formats.

Data scientific thinking

For students to be active participants in learning from modern data and about the modern world, they need to be able to integrate statistical and computational thinking. Developing this kind of *data scientific thinking* does not just happen by using certain computational tools, care needs to be taken to carefully design learning experiences and assessment opportunities that take into account the difficulty of combining statistical and computational understandings. An integrated use of computational approaches also broadens the ways students can learn from data, for example the use of predictive modelling and “discovery” based algorithms, as well as offering the opportunity to use mathematical ideas and representations to respond to contextually-driven purposes.

Statistical and data literacy

There is an urgent need for greater statistical literacy in the age of fake news, media claims/graphics and a proliferation of different types of visualisations. Although statistical literacy is already a strong feature of the New Zealand Curriculum for Statistics, the rest of the world is taking more notice due to the advent of COVID-19 and many are calling for greater teaching of data literacy and graphicacy (see section 1). At the same time, it is important to keep in mind the curriculum expectations for each level in terms of what statistical concepts and representations are appropriate for students to consider. Data visualisations, for example, could be used at all levels of the curriculum, and a clear progression is needed to explain what is appropriate for what levels, including the use of “worry” questions.

Data ethics, responsibilities and ownership

As students are encouraged to interact more closely with data from a wider range of sources, it is imperative that issues such as data ethics, responsibilities and ownership are more clearly present in the teaching and assessment of statistics and data science education[7]. Frameworks such Tertiary Education

---

[6] See https://www.nzqa.govt.nz/assets/About-us/Future-State/NCEA-Online/research-and-innovation/Digital-Stats-2020-Innov-Trial-Eval-Closure-Report.pdf

[7] See paragraphs six to nine of https://oag.parliament.nz/blog/2019/people-and-data

Commission's Ōritetanga learner analytics ethics framework[8] could be used to inform the development of a school-level appropriate approach to teaching data ethics, responsibilities and ownership. A focus on asking about the "whakapapa of your data" provides just one of many ways to integrate Te Ao Māori perspectives and to draw on the mahi and expertise of Māori scientists and researchers, for example, the principles of Māori data sovereignty developed by Te Mana Raraunga[9].

3. What matters the most for statistics and data science education at the senior secondary level in Aotearoa New Zealand?

Making connections
At their core, both statistics and data science education are about facilitating, in developmentally appropriate ways, students' making connections between questions about real-world contexts; sourcing and wrangling data; data exploration and analysis; conjectured, or concluded, generalisations from the data about the context; and understanding and using concepts and tools that enable these things to occur.

Activating statistical dispositions[10]
Ways of learning statistics and data science should include fostering the following:

- *Creativity*
  Examples include: exploring different ways to create something and being open to new ideas; using exploration and visualisation to surface *potential* patterns and relationships (data-informed conjectures as opposed to formal conclusions); pursuing "what if?" questions
- *Scepticism*
  Examples include: using different knowledge systems to question, evaluate & challenge ideas and claims; raising relevant practical or ethical concerns
- *Communication*
  Examples include: presenting ideas clearly, cohesively, convincingly & compellingly; using visualisations, symbols, text, language, orally etc.

Providing opportunities to demonstrate key ideas
Ways of learning statistics and data science should help students demonstrate their understanding of key ideas (knowledge base):

- *Data knowledge*
  This could include: a broad appreciation of the nature of data, social and cultural issues pertaining to data (e.g., ethics, privacy and ownership); other human factors and biases; issues of representativeness and those related to study design
- *Statistical knowledge*
  This could include: key statistical concepts and calculations; features of different representations

---

[8] https://tec.govt.nz/assets/Oritetanga/Learner-Analytics/Analysing-Student-Data-Oritetanga-Learner-Analytics-Framework-updated-June-2021.pdf

[9] https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5bda208b4ae237cd89ee16e9/1541021836126/TMR+Ma%CC%84ori+Data+Sovereignty+Principles+Oct+2018.pdf

[10] Drawing on the dispositions discussed by Wild & Pfannkuch (1999) and Wickham (2010)

and distributions; graphical, tabular and common forms of statistical language (e.g., risk communication, probability notation and manipulation, etc.)

- *Modelling perspective*
  This could include: modelling for structure and variation; different approaches for different goals (e.g. prediction vs explanation); different concepts of probability (e.g. true vs model); accounting for uncertainty; use of random variables; using simulation-based inference; considering model-based assumptions
- *Data scientific thinking*
  This could include: integrating statistical and computational thinking; combining different computational tools; combining different digital sources of data; considering automation and role of technology to scale up data analysis; providing reproducible computational essays

See Appendix A for a more comprehensive list of key ideas for statistics and data science arranged from a more general perspective.

4. Suggestions for key statistics and data science learning experiences for senior secondary level students

Curriculum level seven

- *Explore data using a range of software-enabled visualisations*
  For example: source data relevant to a stated purpose; identify potential patterns and relationships; consider relevant contextual knowledge; communicate key findings; use exploration to inform a focused investigation
- *Use informal methods to make sample-to-population inferences*
  For example: consider data ethics; obtain random samples; generate/visualise informal confidence intervals or comparison intervals[11]; interpret confidence intervals with respect to population parameters
- *Investigate a simple experiment design*
  For example: design a single proportion-based experiment (binomial); use a simulation-based approach to generate model data; compare data collected from an experiment (non-model generated process) to data generated from models; reason with chance variation to reach conclusions
- *Calculate and interpret model probabilities and expectations*
  For example: use representations such as probability tree diagrams and two-way tables to determine probabilities of events; use distributions to model likely/unlikely values; conduct technology-supported simulations to estimate model probabilities; calculate and interpret expected values for simple models (applying expectation algebra)[12]
- *Use a probability model to explore random variation*
  For example: create a probability model; generate outcomes (or products) from a probability or probability distribution model; tinker with the parameters of the model and discuss their effects;

[11] Includes for proportions or differences of two proportions
[12] Here we have used ideas currently taught/assessed at CL8. This is because we are suggesting a focus on the "mathematics" of probability for CL7 e.g. model-world only development of probabilistic thinking.

discuss the relationship between the structural components of the model and the generated outcomes

- *Demonstrate statistical knowledge*
  For example: demonstrate understanding of sampling variation; critique the language used by others when communicating statistical ideas; evaluate claims based on data or models; perform calculations to verify claims based on risks

Curriculum level eight

- *Create visual-based communications by manipulating and merging more than one source of data*
  For example: use sources such as text, images, and audio; explore different ways to encode data; produce effective visualisations; provide a cohesive narrative based on key features of visualisations
- *Use formal methods to make inferences about population parameters*
  For example: construct confidence intervals[13] from sample data using simulation-based methods such as bootstrapping; interpret confidence intervals in context to reach conclusions or evaluate claims
- *Design and conduct experiments for the comparison of two independent groups*
  For example: consider data ethics; apply relevant experimental design principles; use simulation based-methods such as the randomisation test to evaluate strength of evidence; consider study design and implementation when making causal claims
- *Critically evaluate statistically-based communications*
  For example: use rules-of-thumb for estimating margin of error; consider implications of study design for claims (e.g. observational vs experiment); discuss nonsampling errors and biases; interpret complex communications of risk
- *Develop and evaluate predictive models*
  For example: explore data to inform feature selection; use an appropriate method to fit a predictive model; evaluate a model in terms of predictive accuracy; consider generalisability, transferability, overfitting and underfitting
- *Explore probability distribution models*
  For example: understand key features and conditions of probability distributions; use probability distribution models to calculate probability estimates; link parameters of probability distributions to the distribution of probabilities (e.g. shape); use informal tests to assess goodness-of-fit
- *Analyse data using algorithmic approaches*
  For example: use an automated approach to tidy data; discover patterns in data using algorithms such as clustering; develop new data structures or convert existing data structures to another form; create new variables or measures to advance analysis (e.g. measures of similarity)
- *Create and interpret representations of dynamic data or complex systems*
  For example: explore spatial mapping techniques to demonstrate relationships between different variables; develop and visualise a network for a complex system to identify potential relationships between entities; track dynamic time series data and develop interactive reports (e.g. dashboards) that include predictions

[13] Includes proportions or differences of two proportions

5. Closing remarks regarding assessment

Design standards that assess the most important achievement objectives
The current achievement objectives for statistics at levels seven and eight of the New Zealand Curriculum offer a wide range of possibilities for teaching and learning at the senior secondary level. However, not everything described or elaborated on (for example in the senior secondary guide) has to be assessed formally through achievement standards at both levels. Students may not be assessed against all statistics achievement standards available for a particular NCEA level, nor be assessed against statistics achievement standards for both NCEA levels. Therefore, learning pathways, such as sample-to-population inference, should only be embedded in achievement standards at both levels if the learning they represent needs to be covered for two successive years.

Assess only the mathematics that is needed to support statistics
Statistics and mathematics are related, but they use different ways of thinking and solving problems[14]. Both equip students with effective means for investigating, interpreting, explaining, and making sense of the world. The statistics and data science methods taught at the senior secondary level will involve mathematics but these mathematical ideas may not necessarily be at the same level of the curriculum as the statistical ideas being assessed. Hence, in our discussion earlier, we have not considered achievement objectives from the mathematics strands.

Seize the opportunity to overcome shortcomings
Section 2.2 outlined elements of our current assessment standards that have not served us well. We highlighted too little assessment of conceptual understanding, the need to reduce the amount of writing to more manageable levels and for expectations for communication (written or otherwise) to better prioritise quality and clarity of statistical thinking over quantity. Unforeseen consequences of earlier choices also include unnecessarily severe constraints on the types of data that can be used, an absence of any real exposure to exploratory data analysis and a paucity of opportunities to experience many small curiosity-driven explorations. Probability is everywhere, commonly experienced in terms of risk management, and we need to make sure that students who will be our future workforce, decision-makers, and voters have a firm basis in reasoning with uncertainty. It is particularly important that no choices are made that unnecessarily hamper the ability to use modern online data sources that are of high interest to students. We wish the teams working on the new replacement standards all the very best for their endeavours to find creative ways to overcome as many of these as possible!

References

Bargagliotti, A., & Franklin, C. (2015). The statistical education of teachers: Preparing teachers to teach statistics. *Chance*, *28*(3), 19-27.

---

[14] See Section 1 of "*What Is Statistics*?" co-authored by the President of the American Statistical Association and Chair of the Section on Statistics and Data Science Education [Wild, C.J., Utts, J.M., Horton, N.J.. (2018). *What Is Statistics*?. In D. Ben-Zvi., K. Makar & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 5-36). Springer, Cham. Preprint of Section 1]

Ben-Zvi., D., Makar, K. & Garfield, J. (Eds.) (2018), *International Handbook of Research in Statistics Education*. Springer, Cham. https://link.springer.com/book/10.1007/978-3-319-66195-7

Biehler, R., & Fleischer, Y. (2021). Introducing students to machine learning with decision trees using CODAP and Jupyter Notebooks. *Teaching Statistics*, *43*, S133-S142.

Case, C., & Jacobbe, T. (2018). A framework to characterize student difficulties in learning inference from a simulation-based approach. *Statistics Education Research Journal*, *17*(2), 9-29.

Cook, D., Reid, N., & Tanaka, E. (2021). The Foundation is Available for Thinking about Data Visualization Inferentially.

Engel, J. (2019). Statistical literacy and society. What is civic statistics. In *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* (Vol. 39, No. 1, pp. 1-12).

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*(2), 739-762.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22-25.

Gould, R. (2021). Toward data‐scientific thinking. *Teaching Statistics*, *43*, S11-S22.

Jones, R. C. (2019). Statistical investigation measuring intelligence and creativity. *Teaching Statistics, 41* (1), 36-40.

Kazak, S., & Pratt, D. (2021). Developing the role of modelling in the teaching and learning of probability. *Research in Mathematics Education*, *23*(2), 113-133.

Pfannkuch, M., Wild, C., Arnold, P., & Budgett, S. (2020). Reflections on a 20-Year Statistics Education Research Journey, 1999-2019. set: Research Information for Teachers, 1, 27-33.

Tintle, N., Clark, J., Fischer, K., Chance, B., Cobb, G., Roy, S., ... & VanderStoep, J. (2018). Assessing the association between precourse metrics of student preparation and student performance in introductory statistics: Results from early data on simulation-based inference vs. non simulation-based inference. Journal of Statistics Education, 26(2), 103-109.

Wickham, H. (2010). Using visualisation in teaching data analysis and programming. In *Invited paper presented to the Eighth International Conference on Teaching Statistics (ICOTS 8).*

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review*, *67*(3), 223-248.

APPENDIX A: *Summary of a statement submitted to the Royal Society Te Apārangi panel convened to provide advice to support the refresh of 'The New Zealand Curriculum' and 'Te Marautanga o Aotearoa' and to inform the development of a high-level systems approach to supporting learning in mathematics.*

Introduction

Statistics is defined by the American Statistical Association (ASA) as "the science of learning from data, and of measuring, controlling and communicating uncertainty."

The first decades of the 21st century has heralded an unprecedented data revolution fundamentally impacting our daily lives, often in ways we would never have dreamed possible – a world in which the business models of some of the five richest corporations in the world are based on monetising personal data (with huge downstream impacts on society). Having already heavily embraced technology, statistics education continues to evolve in response to further rapid changes in technology, and in changes in the statistics discipline, how society interacts with data-based evidence, and to the continual emergence of new types and sources of data, e.g., digital-tracking data. We are in a world that is increasingly data centric, where the nature of data and ways of handling data are rapidly changing, and where statistics is moving beyond traditional inferential emphases to computationally intensive approaches and new ways of processing and reasoning with data. We need to prepare students for this rapidly evolving world.

Statistics is important because there is a "fundamental human need to be able to learn about how our world operates using data, all the while acknowledging sources and levels of uncertainty." Statistical methods are used in almost all knowledge areas and are increasingly used by businesses, governments, health practitioners, other professionals, and individuals to make better predictions and decisions. Conclusions and advice based on statistical methods abound in the media. Some of the thinking used for decision-making based on quantitative data carries over into decision-making involving uncertainty in daily life even when quantitative data are not available.

The last decade has seen unprecedented growth in the availability of data in most areas of human endeavour. Whole branches of science have been developed to allow corporations to transform the way marketing is conducted, and to drive scientific progress in areas such as bioinformatics, and to inform decision-making at all levels in governments and industry. Further, the scale and complexity of much of these data are beyond the capability of single computers to manage or a single individual to analyse.

In preparing our students to participate in this data-driven world, all levels of schooling will need to address the development of students' current and future skills, knowledge and conceptual understandings.

The important "big ideas" in statistics that all learners need to develop through schooling are:

- Data is needed for learning about the world, situations, or phenomena, and for making predictions, decisions and evaluating risks.
- Scrutinising the credibility of data-based evidence, conclusions, and decisions is becoming an essential requirement for citizens to fully participate in a democracy.
    - Statistical inquiry is about obtaining data to answer real-world questions and unlocking the stories in that data.

    - Variation, uncertainty, and random behaviour occur everywhere and need to be taken into account in all phases of statistical investigations and in the communication of findings.
    - Distribution as the lens through which variation in data is viewed and comprehended.
    - Visualising data enables variation, central tendencies, patterns, and exceptions to be seen.
- Simulations using empirical and model data, as well as real-time and big data, help in the understanding the behaviour of real-world systems and in making predictions (well-known post COVID-19!).
- Randomness and probability help us model unexplained variation produced by real systems (components of stochastic models).
- Data usually has some social context and impact. We should worry about the social and ethical dimensions of the data we use, create, manage, or disseminate and think very carefully about how we should act.

To become a **critically engaged citizen the statistical skills and knowledge needed** are many-fold to participate meaningfully in society.

- First, students need to realise that data is needed to learn more about the world, to be able to judge a situation and to make an informed decision, and that evidence from good data should always outweigh personal stories (anecdotal evidence).
- Second, students need to develop and use a set of "worry questions" when faced with data-based evidence in the media, in their everyday lives, for example, regarding decisions about health, finance, lifestyle issues, and in society whether as a member of a community at the local or national level. These worry questions include where the data come from, why were the data collected, how the data were created and how measures were defined, whether there could be an alternative explanations for the findings, whether generalisations are reasonable, and whether one's beliefs may subconsciously prevent paying attention to evidence that does not align with one's world view. Paying attention to worry questions can allow students to form and communicate a reasoned opinion on whether the findings are trustworthy and help habituate critical thinking.
- Third, students need to become more aware about how data is being collected about them and how this data is used to train algorithms and to develop models. They need to question whether the training of algorithms for prediction are biased in favour of the status quo and may be treating specific groups of people unfairly, whether uncertainty was taken into account, and what steps were take to ensure that no harm is caused when these algorithms are put into production. Students also need experience formulating and testing data-driven algorithms to understand the big ideas well enough to be able to think critically about them, and to appreciate potential difficulties with artificial-intelligence solutions.
- Fourth, access to publicly available data has seen the rise of Citizen Statistics (e.g., the Global Indigenous Data Alliance) where the public and students can not only engage in the data collection process but also the analysis and interpretation of data for the purpose of advocacy as a citizen, in order to improve outcomes.
- Fifth, as Te Mana Raraunga Māori Data Sovereignty Network states, data ethics, privacy, confidentiality and ownership are paramount considerations for a functioning democracy and therefore students need to appreciate these issues.

Statistics is inherently **cross-disciplinary**, as data are ubiquitous in many fields of endeavour and hence in many school curriculum areas such as science, social science, history and technology. Contextual knowledge, that is, knowledge of the subject area from which the data arise, is essential for interrogating and understanding data, with the ultimate goal of statistical investigation being learning more about the context sphere (WP99). With the **rapid change and growth in computer intensive methods**, statistics education needs to continually renew and adapt to new ways of interacting and engaging with data. Students, as novice statisticians, need to engage in statistical practices and thinking that emulates current professional discipline through age-appropriate learning approaches. Conducting statistical investigations involves polls and surveys, observational studies and randomised controlled experiments, which are at the traditional core of the statisticians' art, or the mining of sources of existing data (which brings a different set of challenges for generalisation). The enculturation and induction of students into these core practices is important as well as the development of skills, knowledge and concepts associated with statistical reasoning.

Traditionally, during curriculum development, subjects that require statistical skills are developed independently with little or no regard to the statistics skills and knowledge that students have in their toolbox to apply in an interdisciplinary way. Breaking down subject silos at the curriculum development level would help students to transfer knowledge and skills across the disciplines and reduce confusion.

Statistics crosses almost all the disciplines because it is the science of learning about the real world from data. Statistical thinking enriches students' understanding of the art of: posing questions (investigative, survey, analysis, and interrogative), noticing and wondering, tinkering, actions and reasoning, the data detective unlocking and telling stories revealed in the data, data creation, planning for and collecting data, considering and using existing data sets, accounting for variation, structuring and representing data in graphical and tabular distributions to perceive variation, central tendencies, patterns and exceptions, reasoning from and about graphs, infographics and tables with context as integral for meaning-making, probabilistic reasoning, interpreting, arguing with data-based evidence, justifying and communicating findings and making decisions under uncertainty. Essential also, is students becoming familiar with statistical modelling through conducting simulations using empirical and model data as well as experiencing real-time and big data.

Purpose-built educational software (e.g., CODAP, iNZight, TinkerPlots, VIT) that forefronts the development of students' statistical conceptual understanding as well as providing data analysis tools should be used at all levels of schooling. Interactive visualisations and visual methods of learning statistics and statistical and probabilistic reasoning are essential ingredients towards learning how to handle and deal with data. Consideration needs to be given to how students can build on their statistical knowledge and skills to incorporate computational ways of thinking. For example, the use of coding as part of statistical investigations needs to be implemented in a way that supports students to think statistically with data rather than to prioritise precision or efficiency with processing data.

Numeracy is foundational to statistics yet there is a strong synergy between them. As well as being a core element of numeracy, the interpretation of numbers in context is entirely core to statistical understanding and communication. As students develop statistical ideas, numeracy ideas can be co-constructed and/or reinforced. Statistics provides a context for the number and measurement content ideas of the proposed numeracy standard as well as being part of the numeracy standard. The synergy can be realised by developing student understanding, as statistics provides a rationale, for example, to measure height for data collection, to use percentages to compare the categorical outcomes of two groups with different sample sizes, or the risks between several groups. However, there is a danger in treating probability as a deterministic application of fractions, decimals or percentages because developing probabilistic reasoning involves dealing with uncertainty, randomness and the omnipresence of variation. Proportional reasoning underpins statistical reasoning as students grapple with interpreting and reasoning from data. Interpreting percentages and decimals are inherent in probabilistic reasoning (e.g., what is the probability that I have a disease given that I have a positive test). Understanding relative risk, absolute risk, and increased risk are foundational conceptions for producing risk savvy citizens and involve fractions, decimals and percentages. Hence, the relationship between numeracy and statistics is inextricably intertwined.

Data science provides an opportunity to engage students with digital technologies as part of explorations and investigations with modern data. As stated earlier, with the rapid change and growth in computer intensive methods, statistics education needs to continually renew and adapt to new ways of interacting and engaging with data. Key areas of data science that have yet to be realised at the school level include: predictive modelling, sourcing data in dynamic ways for investigation including the use of APIs or web scraping, generative modelling, data visualisation. To participate in data science requires careful consideration of how to integrate statistical and computational thinking to learn from data. This necessitates providing age-appropriate learning opportunities for students across all year levels to:

- Recognise and measure data from a wide range of digital sources, including image, text, spatial, sound, and human interactions through networks such as social media.
- Engage with open data-based explorations to generate ideas for future model-based solutions.
- Work with complex and "messy" data sets for the purpose of drawing insights about the data-context.
- Learn techniques for data extraction and manipulation, as part of collecting and analysing data across a range of data sources, files and structures.
- Use a wide range of graphical techniques to communicate visual-based arguments, selecting appropriate technology and software.
- Use mathematical representations and computational techniques in the process of developing models, rules and generalisations.
- Implement and evaluate algorithmic modelling approaches.
- Reason critically with data, models and visualisations when forming arguments or making decisions.
- Produce reproducible digital reports that integrate statistical and computational thinking and clearly communicate uncertainty and methods transparently.
- Describe and apply responsible and ethical practices when obtaining and using data from public sources
- Consider practical and social consequences of data-based decisions.
- Apply computation (including coding) to support their learning from data.

Data science combines elements of knowledge and thinking from statistics, mathematics and computer science (digital technologies), linked together by the same purpose of providing a way to learn about the context of the data and to create products or solutions for that data-context. However, currently none of these learning areas by themselves provide the learning experiences just described. Furthermore, urgent research is needed to inform the development of curriculum and learning tasks that will engage all students with data science across all year levels, in ways that will ensure equity and reduce barriers to participation.

APPENDIX B: *Summary of our statement to the Ministry of Education about the proposed NCEA Level 2 and 3 subjects.*

We strongly support the creation of the subjects Mathematics and Statistics at Levels 2 and 3
The Ministry asks whether *'the proposed split into two subjects, Mathematics and Statistics, at Level 2, covers all significant learning that should be available'.* One subject (statistics) can cover the statistics part of the learning, if the four Achievement Standards are carefully designed to incorporate the latest understandings of how the subject can best be taught, learned, and assessed at that level. We look forward to working with the Ministry to achieve this.

The same applies to the creation of the two subjects, Mathematics and Statistics, at Level 3. Both subjects, Mathematics and Statistics, are vital for the future of New Zealand. At both levels, we would like to see a design of the eight Standards that allows ākonga to build a firm foundation of the essential learning.

We see a great opportunity in the third L3 subject for a transformational step into the future
The Ministry asks whether *'the proposed split into three subjects, Mathematics, Statistics and Applied Mathematics, at Level 3 covers all significant learning that should be available',* and asks for *'other comments about Mathematics and Statistics; for example, what learning do you think should be included in the new Applied Mathematics subject to help ākonga prepare for life, work, and further study?'*

We are happy to see the number of credits available for Level 3 Mathematics and Statistics remains at 60 credits and that a new third Level 3 subject has been proposed. This new subject could allow kura to develop and strengthen options and pathways for ākonga, and to create meaningful and stimulating courses that weave mathematical and statistical ideas within personal and cultural contexts. We believe that to help ākonga prepare for life, work, and further study, it is vitally important to provide a subject that embraces technology and focuses on *integrating* statistical, computational, and mathematical thinking to learn from data-driven contexts through software-enabled modelling and visualisation.

In past decades, Aotearoa New Zealand has made a series of decisions about extra 'mathematics' which were all sensibly aimed at the applications for the mathematics of the day: mechanics, numerical methods, and the operational research methods currently taught. It is time for another sensible and bold decision, in this rapidly changing world. We need a reset and not a recycle. We have the chance now to do something that is exciting, new, relevant to today's world as well as future-facing. At this moment, Aotearoa New Zealand has a golden opportunity to create a future-focused subject in a co-ordinated and nationally supported manner.

This new subject should enable ākonga to flourish in the rapidly changing world of data technologies, equip them with a broad base of computationally-driven modelling capabilities and connect to the world they live in in ways that will be obvious to ākonga. We see such expertise and enthusiasm in the country's teaching community, and we would like Aotearoa New Zealand to keep its world-leading position in curriculum development by embracing data science as the basis of the third Level 3 subject for the mathematics and statistics learning area.

This new Level 3 subject needs a cohesive and future-focused identity

The Ministry describes the proposed new Level 3 subject as one where "*ākonga will engage with Mathematics and Statistics content (with some overlaps with both Mathematics and Statistics), but with an emphasis on the applied nature of the subject and how it relates to other subjects*". We also want to see an applied focus for this new subject, one where ākonga integrate statistical, computational and mathematical thinking to learn from data-driven contexts through software-enabled modelling and visualisation.

We would like to see a subject that has as much breadth and academic challenge as mathematics and statistics but that is grounded on a modelling approach that utilises modern software and technology (including coding). This new Level 3 subject should be empowering and relevant for all students, including those who may also study Level 3 Mathematics or Statistics concurrently. The reason we have stated that data science should be the foundation of this new Level 3 subject is because it could offer the kinds of fully blended learning experiences for ākonga we believe the Ministry envisages.

The discussion document proposes that topics like networks, linear programming, and logic will form the basis of the new Level 3 subject. The technical report further proposes truth tables, synthetic languages and algebra. The suggested topics contain no statistical content at all. But even if we set that aside, what is suggested is a list of disconnected, specialised, technical topics that conveys no vision of applied mathematics in its real sense of applying mathematical techniques to solve real problems in the world outside of mathematics. A smorgasbord of such topics conveys no sense of the power of mathematical and statistical techniques to illuminate the real world in which modern ākonga live. They are optional, "on the side" topics that are not required for standard pathways – or in the case of algebra a core part of *pure* mathematics. While some may help a little with a few specialities for some future studies they are not important future-facing topics "to prepare a broad spectrum of ākonga for life, (and) work".

The discussion document also clearly states that a subject must represent a separate and distinct body of knowledge rather than a collection of topics or contexts of another subject. The four achievement standards that are developed for this new subject must be complementary and be clearly identifiable as that subject. Furthermore, a cohesive 21st century subject needs to be defined based on key ideas, practices and pedagogies, with transferable skills and knowledge for the modern world. The fact that a third subject appears for the first time at curriculum level eight allows for a subject to stand alone alongside mathematics, statistics, and digital technology, unfettered by their marches towards expertise in calculus, inferential statistics and computer-programming respectively. As such, it would also provide an opportunity to re-engage students who have become disengaged from mathematics and statistics (see section 5).

Why should we introduce a new subject based on data science?

There is a perception that data science could be taught in a school course by offering topic areas and subsequent assessments from within the suite of statistics, mathematics, and digital technologies achievement standards. However, this does not address the issue that targeted teaching and learning experiences are needed for ākonga to learn how to integrate statistical, computational, and mathematical thinking throughout the process of modelling. Combining new ideas from across a range of subjects, particularly when digital technologies are involved such as coding, requires specific pedagogies to ensure all ākonga are included and supported in their learning. Existing achievement standards would not provide appropriate assessment options for the course we envision. It is unlikely future achievement standards from these domains would do so either, simply because the thinking modes and learning contexts are unique to data science and the modelling-based learning environment that it encompasses.

Our experiences with teaching data science approaches is that ākonga cannot be expected to blend elements for themselves without carefully designed learning experiences that focus on inclusion for all ākonga and use relevant and meaningful social and cultural contexts. For example, we have seen ākonga who did not believe they belonged in mathematics and statistics thrive in a computationally-based learning environment that carefully introduced code-driven tools to support creativity with modelling. Therefore, it is important that the different thinking modes and learning contexts that are needed to support ākonga to actively participate in a subject are recognised and promoted by its own specific Achievement Standards. Introducing a new subject at Level 3 must be mindful of the streaming-by-default practices that take place in many schools, in particular the use of “pre-requisites”. We would like to see this new third subject taught in such a way that ākonga can be successful, regardless of their past experiences in mathematics and statistics.

We are happy to engage with the Ministry on some options for the naming of the third Level 3 subject but do not believe that ‘Applied Mathematics’ is appropriate for either the current proposal or a subject based on data science. The suggested name 'Applied Mathematics' of the current proposed course has historical baggage at high school level, often being the name for lower-level courses. Additionally, to name the subject using ‘Mathematics’ without reference to Statistics does not make it clear that ākonga will engage with both Mathematics and Statistics ideas.

We acknowledge that ‘Data Science’ may also suffer from the same critique, if ‘data’ and ‘data science’ are viewed as only being associated with Statistics - which is not how these terms are associated in business, industry and government. However, we believe a new name would clearly signal a new subject, and also encourage a fresh approach to designing the subject’s identity. Additionally, there is an international movement for data science to be taught at the high school level and other countries have already adopted high school data science curricula. Data science meets the criteria for a subject that is “fit for purpose for 21st -century learners” and reflects “emerging developments in education and the world of work.”

Implementing data science at the high school level in Aotearoa

Data science is a growing field, with many major universities now offering data science majors both internationally and within Aotearoa New Zealand. Data science curricula for the high school level are being developed and taught around the world. Some of these include: the International Data Science in Schools Project (IDSSP) (http://www.idssp.org/), the US-based Introduction to Data Science curriculum (IDS), the Germany-based ProDaBi curriculum (https://www.prodabi.de/), and the Youcubed Explorations in Data Science curriculum (https://hsdatascience.youcubed.org/).

Aotearoa New Zealand has an opportunity to be world-leading, to be bold, and to develop an enviable vision for data science, one that integrates te ao Māori and mātauranga Māori from its inception and recognises the unique qualities of our current curriculum and of our country. We must ensure the knowledge, experiences and expertise of Māori data scientists and others melding te ao Māori and western approaches in the field of data science are not just valued but drive the development of this subject. We have the opportunity to define what data science education could look like for our ākonga, and we must weave together diverse cultural perspectives on what's important for this new Level 3 subject.

Data science can be taught independently of any other subjects. Integrating statistical, computational, and mathematical thinking to learn from data-driven contexts through software-enabled modelling and visualisation could involve elements like:

- Describing and applying responsible and ethical practices when obtaining and using data from digital sources
- Accessing data from modern data sources and contexts such as social media, GPS, APIs, web scraping, sensors, and other digital mechanisms
- Manipulating, reshaping, and merging data from different sources and structures, including digital image and audio data, text analysis, and spatial temporal data
- Using mathematical representations and computational techniques in the process of developing models, rules and generalisations
- Creating effective and reproducible visual-based communications, including automated reporting systems (e.g. business reports) or interactive graphics such as dashboard
- Using supervised machine learning techniques to build and test predictive and classification models
- Exploring ideas of optimisation using algorithms such as in clustering
- Modelling some complex systems (e.g. spatial systems, networks or processes evolving over time)
- Considering social and practical consequences of modelling decisions and algorithmic biases

The mathematics and computation that underlies some of these activities can be described 'along the way' or as post hoc explanations of some 'under the hood' things that have empowered ākonga to do what they have just done. Assessment activities should not be based on demonstrating specific, abstract, or algorithmic mathematical, computational or statistical ideas "by hand". Instead, they should encourage ākonga to integrate statistical, computational, and mathematical thinking to learn from data-driven contexts through software-enabled modelling and visualisation. We propose that a software-enabled data science subject could do all of the above, feeding off meaningful engagement with modern data sources providing for some expansive big-picture engagement to complement the more detailed-level experiences that form the bulk of the education of ākonga in mathematics and statistics.

Doing anything new and future-facing requires substantial teacher re-education. Introducing the third subject within the mathematics and statistics curriculum area enables growth over time from a small base, rather than needing the initial mass re-education of a whole cohort of teachers. By defining a new subject based on data science we have an opportunity to have a co-ordinated plan for pedagogical and resource development and upskilling of teachers. The development of this new third subject needs to pull together international research and best practice viewed through the lens of Aotearoa New Zealand. It would be crucial to work with the tertiary sector, as they have also been developing their data science curricula and can share expertise, best practice, and research to support data science at the high school level. We are incredibly fortunate in Aotearoa New Zealand that we have such experts willing to support our high school level teachers. With piloting due in 2024 and implementation of new subjects and standards in 2025, the time frame should not be seen as a barrier.

The new subject should be designed with student well-being and equity at its heart

Historically, mathematics has served as a gatekeeper subject, with access to higher level courses dependent on success in lower-level courses. Although ākonga will benefit from having studied mathematics or statistics before, it is important that the third subject at Level 3 enables success for all students, no matter their past experiences in mathematics and statistics. This new third subject should be used as an opportunity to re-engage ākonga who may have not been well served by our current education system. We could and should use such a restart to reach out to wider audiences, and also re-engage and re-energise those who have become disengaged from the existing pipelines. With this new subject, we can directly address issues of equity, de-streaming, and valuing broader concepts of 'knowledge'. The ākonga who we imagine taking the subject that is currently called 'Applied Mathematics' have futures as important as ākonga taking any other subject at NCEA Level 3. Here is our opportunity to create a future-focused, modern, inclusive, and empowering course that will prepare them for life and work as well as support tertiary pathways.

APPENDIX C: *Some recent journal publications or book chapters by NZ statistics and data science education researchers involving NZ students or teachers, NZ curriculum or NZ-developed educational technology, grouped by theme/topic.*

General

Pfannkuch, M., Wild, C., Arnold, P., & Budgett, S. (2020). Reflections on a 20-Year Statistics Education Research Journey, 1999-2019. *set: Research Information for Teachers*, *1*, 27-33.

Statistical investigations

Arnold, P., & Pfannkuch, M. (2019). Posing comparative statistical investigative questions. In *Topics and trends in current statistics education research* (pp. 173-195). Springer, Cham.

Arnold, P., & Pfannkuch, M. (2020). On Being Data Detectives: Developing Novice Statisticians Using the Statistical Enquiry Cycle. *set: Research Information for Teachers*, *1*, 34-41.

Probability or probability modelling

Budgett, S., & Pfannkuch, M. (2019). Visualizing chance: tackling conditional probability misconceptions. In *Topics and Trends in Current Statistics Education Research* (pp. 3-25). Springer, Cham.

Daly, N., & Sharma, S. (2018). Language-as-Resource: Language strategies used by New Zealand teachers working in an international multilingual setting. *Australian Journal of Teacher Education* (Online), 43(8), 15-29.

Dayal, H.C. & Sharma, S. (2020) Investigating probability concepts of secondary pre-service teachers in a game context. *Australian Journal of Teacher Education* (Online) 45 (5), 91-109.

Fergusson, A., & Pfannkuch, M. (2020). Development of an informal test for the fit of a probability distribution model for teaching. *Journal of Statistics Education*, *28*(3), 344-357.

Patel, A. & Pfannkuch, M. (2018). Developing a statistical modeling framework to characterize Year 7 students' reasoning. *ZDM, (50)*7, 1197-1212.

Renelle, A., Budgett, S., & Jones, R. (2021). New Zealand teachers' generation problem misconceptions. *Teaching Statistics*, *43*(2), 56-61.

Sharma, S. (2019), Language Challenges and Strategies for English Language Learners in Statistics Education: An Overview of Research in This Field. *Education Quarterly Reviews*, Vol.2, No.3, 651-665.

Data science

Burr, W., Chevalier, F., Collins, C., Gibbs, A. L., Ng, R., & Wild, C. J. (2021). Computational skills by stealth in introductory data science teaching. *Teaching Statistics*, 43, S34-S51.

Fergusson, A., & Pfannkuch, M. (2021). Introducing teachers who use GUI-driven tools for the randomization test to code-driven tools. *Mathematical Thinking and Learning*, 1-21.

Fergusson, A., & Wild, C. J. (2021). On traversing the data landscape: Introducing APIs to data-science students. *Teaching Statistics*, *43*, S71-S83.

Wild, C. J., Elliott, T., & Sporle, A. (2021). On Democratizing Data Science: Some iNZights Into Empowering the Many. *Harvard Data Science Review*.

APPENDIX D: *Some recent collaborations between NZ statistics and data science education researchers and international educators, researchers or educational organisations.*

Guidelines for Assessment and Instruction in Statistics Education II (Anna Bargagliotti & Christine Franklin - co-chairs - with Pip Arnold, Rob Gould, Sheri Johnson, Leticia Perez, Denise Spangler)

Arnold, P., & Franklin, C. (2021). What Makes a Good Statistical Question?. *Journal of Statistics and Data Science Education*, *29*(1), 122-130.

Bargagliotti, A., Arnold, P., & Franklin, C. (2021). GAISE II: Bringing Data into Classrooms. *Mathematics Teacher: Learning and Teaching PK-12*, *114*(6), 424-435.

International Handbook for Statistics Education (Editors: Dani Ben-Zvi, Katie Makar, & Joan Garfield)

Arnold, P., Confrey, J., Jones, R. S., Lee, H. S., & Pfannkuch, M. (2018). Statistics learning trajectories. In *International handbook of research in statistics education* (pp. 295-326). Springer, Cham.

Pfannkuch, M. (2018). Reimagining curriculum approaches. In *International handbook of research in statistics education* (pp. 387-413). Springer, Cham.

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics?. In *International handbook of research in statistics education* (pp. 5-36). Springer, Cham.

International conference for Statistical Reasoning Teaching and Learning (Organisers: Katie Makar & Dani Ben-Zvi)

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM*, *50*(7), 1113-1123.

International Data Science in Schools Project: Chris WIld (lead curriculum author) - with Anna Fergusson, Michelle Dalrymple (Advisors)

IDSSP Curriculum Team (2019). Curriculum Frameworks for Introductory Data Science. http://www.idssp.org/pages/framework.html.

Burr, W., Chevalier, F., Collins, C., Gibbs, A. L., Ng, R., & Wild, C. J. (2021). Computational skills by stealth in introductory data science teaching. *Teaching Statistics*, *43*, S34-S51.

Statistics for empowerment and social engagement: Teaching Civic Statistics to develop informed citizens Chris Wild (chapter) with Jim Ridgway (Editor)

Wild, C. J., & Ridgway, J. (2021). Civic statistics and iNZight: illustrations of some design principles for educational software. *Statistics for empowerment and social engagement—Teaching civic statistics to develop informed citizens. Springer*. Preprint: