TLRI Project Plus+

Nāu i whatu te kākahu, he tāniko taku

Doing research that matters: A success story from statistics education

Rosemary Hipkins

New Zealand Council for Educational Research
PO Box 3237
Wellington
New Zealand

ISBN 978-1-927231-41-8

Contents

Introduction 5

About this TLRI report 5

An introduction to the selected projects 9

The main players 9

Keeping track of the action 10

Doing research that matters 11

Designing convincing investigations of practice 17

Choosing an appropriate methodology 17

Directing design efforts where they will have most impact 17

Gathering clear evidence of the impacts on learning 19

Resources as products of scholarship in practice 21

Creating resources to test new teaching approaches 21

Resources as 'things to think with' 21

Ensuring the products of scholarship are used appropriately 22

Transferring these ideas to other projects 23

Advancing the knowledge and skills of teachers 25

The projects from the participating teachers' perspectives 25

Supporting the professional learning of their peers 26

Leveraging other professional learning opportunities 26

Other benefits for teacher partners 26

Conceptualising relationships between partners 27

Leveraging change at the policy level 29

Leveraging curriculum change 29

Leveraging assessment change 30

Marshalling the support of powerful allies 31

Creating overall conditions for successful knowledge-building in practice 33

The role of the international community of statisticians and statistics educators 33

Other evidence of positive consequences 34

Concluding comments 37

Timelines 38

References 41

Acknowledgements 43

Introduction

About this TLRI report

This is the first report from a new initiative called TLRI Project Plus. The Teaching and Learning Research Initiative (TLRI) provides government-funded support for research that builds new knowledge about teaching and learning in Aotearoa/New Zealand.[1]

Taking a meta-view, TLRI Project Plus will add value to TLRI by synthesising findings across multiple projects. Two projects with a common focus on aspects of students' statistical thinking capabilities provide the foundation for this first TLRI Project Plus report. Based on feedback from the research leaders and others who were involved in one or both projects, this report explores factors that contributed to their success.

The most successful TLRI projects generate new knowledge about teaching and learning, and also have an impact on practice in important ways. While every project is unique, my aim in developing this report has been to inspire strong design work for future projects. I also aim to demonstrate and celebrate the potential of the TLRI model for generating new knowledge that is both practically useful and theoretically robust.

Key design features of TLRI

The TLRI fund aims to build new knowledge about teaching *and* learning. This dual focus points to the importance of unravelling the relationships between intentional acts of teaching and specific instances of learning. TLRI research projects must also be designed as *partnerships*, in which teaching practitioners work alongside researchers. The intention is that a dual focus on teaching and learning, in combination with the partnership requirement, will create conditions that enable relevant bodies of theoretical expertise and deep knowledge of practice to be brought together. In this model, *knowledge building* about teaching and learning is positioned as a collaborative research endeavour.

Adopting a model of collaborative knowledge building has implications for many different facets of any TLRI project. The requirement to work with a research/practice dynamic, and a dual focus on teaching and learning, influences the types of questions that can be posed, the types of methodologies that can be usefully employed, and the types of evidence that can be collected. A partnership model demands that careful attention be paid to the ways in which different types of expertise are deployed at all the different project stages. Clearly, making the most effective use of the knowledge and skills of a diverse team also has implications for overall project leadership. Successfully holding all these complex pieces together could be seen as demanding a specific form of scholarship—scholarship in practice.

The idea of 'scholarship in practice'

The theoretical argument that there is such a thing as *scholarship in practice* is relatively new (Singer-Gabella, 2012). It was developed in response to pressures placed on tertiary educators by accountability measures that tend to privilege traditional forms of research outputs (e.g. publication in high-status journals) while overlooking other types of outputs that might also be important to collaborative research/practice endeavours in education,

1 See http://www.tlri.org.nz/about-tlri for more background information.

such as shaping careful, evidence-informed policy reports or creating significant new approaches to addressing a persistent problem of practice.

The discussion that follows draws on the six criteria proposed by Singer-Gabella that would mark out these types of endeavours as legitimate forms of scholarship. Following are my abbreviated summaries of her six criteria.

- **Compelling conception of ends:** Those involved are clear about how and why addressing the challenges of a specific project could help to generate knowledge that advances practice. Both depth and breadth of relevant expertise are needed to successfully address the ends envisaged.
- **Appropriate means or methods:** The means or methods employed are appropriate to the intended ends and are deployed in an "intentional, skilled, dynamic" manner. The process used is often recursive, as evidence emerges about whether and how the means support the project's ends.
- **Sound judgement:** Scholarship in practice demands that both theoretical and practical knowledge are brought to bear, carefully grounded in the contextual specifics of the challenge being addressed. Choices and their consequences are made visible so that the products of the scholarship can be more readily transferred to other contexts.
- **Significant outcomes:** Scholarship in practice produces local and immediate outcomes that advance the knowledge and skills of teachers and other educators who are directly affected by the work.
- **Generativity:** This criterion shifts the focus to longer-term and less immediate outcomes. The work has impacts beyond the immediate context. It opens up new ideas and directions of enquiry. The work speaks to a wide range of stakeholders: teachers and teacher educators, educational administrators, and education policy makers.
- **Reflective critique:** Clear connections are established between 'reasoning in action' (i.e. judgements made) and the consequences of the work. There is breadth and depth in the evidence presented to support this critique, and significant implications are drawn for future practice.

This report discusses how these six criteria for scholarship in practice have been met within the two selected TLRI projects.

Basically these two projects have resulted in a range of outputs that would not 'count' quite as cogently as evidence of traditional scholarship. For example, the research outputs include new teaching and learning tools and approaches that are now being used by mathematics teachers in a wide range of New Zealand schools. It is these outputs—and specifically the ways of working needed to get to robust outputs of this type—that exemplify scholarship in practice. This less familiar way of thinking about being scholarly when building practical/theoretical knowledge could be helpful for other teams looking to build their profiles as scholars at the demanding interface of research and practice, when the focus is on the inter-relationships between teaching and learning.

However, it is worth noting that these projects have also been highly successful in achieving the robust theoretical outputs of traditional scholarship. They have generated many articles for peer-reviewed journals, and the work has drawn significant international attention. These publications are listed in Box 1. As one of the project leaders noted, such publications are essential to gain attention and credibility in the wider international community of statistics educators.

Box 1: **Scholarly accounts related to the two TLRI investigations**

Arnold, P., Budgett, S., & Pfannkuch, M. (2013). Experiment-to-causation inference: The emergence of new considerations regarding uncertainty. *Proceedings of the 8th International Collaboration for Research on Statistical Reasoning, Thinking and Learning, 18–24 August 2013, Two Harbors, Minnesota, USA* (pp. 119–146). Minnesota University: SRTL-8.

Arnold, P., & Pfannkuch, M. (2010). Enhancing students' inferential reasoning: From hands on to 'movie snapshots'. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society: Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8),July 2010, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications.php

Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to 'movies'. *Journal of Statistics Education, 19*(2), 1–32. Retrieved from http://www.amstat.org/publications/jse/v19n2/pfannkuch.pdf

Budgett, S., & Pfannkuch, M. (2014). Developing statistical inferential concepts in introductory courses. *Proceedings of the 59th International Statistical Institute World Statistical Congress, 25–30 August 2013, Hong Kong, China* (pp. 537–542). The Hague, The Netherlands: International Statistical Institute. Retrieved from http://2013.isiproceedings.org/

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2012). Dynamic visualizations for inference. *International Association for Statistical Education Roundtable Conference: Technology in statistics education: Virtualities and realities, 2–6 July 2012, Cebu City, The Philippines*. Retrieved from http://icots.net/roundtable/programme.php

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education*, *7*(2), 1–21. Retrieved from http://escholarship.org/uc/item/9dg6h7wb

Budgett, S., & Wild, C. J. (accepted). Students' visual reasoning and the randomization test. In K. Makar & R. Gould (Eds.), *Sustainability in statistics education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), July 2014, Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute.

Cumming, J., Miller, C., & Pfannkuch, M. (accepted). Using bootstrap dynamic visualizations in teaching. In K. Makar & R. Gould (Eds.), *Sustainability in statistics education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), July 2014, Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute.

Parsonage, R., Pfannkuch, M., Wild, C. J., & Aloisio, K. (2012). Bootstrapping confidence intervals. *Proceedings of the12th International Congress on Mathematics Education, Topic Study Group 12, 8–15 July, Seoul, Korea* (pp. 2613–2622). Seoul, Korea: ICME-12. Retrieved from http://icme12.org/

Pfannkuch, M. (2010). Inferential reasoning: Learning to 'make a call' in practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society: Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8), July 2010, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications.php

Pfannkuch, M., Arnold, P., & Wild, C. J. (2011). *Building students' inferential reasoning: Statistics curriculum levels 5 and 6. Statistics: It's reasoning, not calculating.* Retrieved from http://www.tlri.org.nz/tlri-research/research-completed/schoolsector/building-students-inferential-reasoning-statistics

Pfannkuch, M., Arnold, P., & Wild, C. J. (2014). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics.* doi: 10.1007/s10649-014-9539-1

Pfannkuch, M., & Budgett, S. (accepted). Constructing inferential concepts through bootstrap and randomization-test simulations: A case study. In K. Makar & R. Gould (Eds.), *Sustainability in statistics education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), July 2014, Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M., Budgett, S., & Arnold, P. (accepted). Experiment-to-causation inference: Understanding causality in a probabilistic setting. In A. Zieffler (Ed.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning.* Minneapolis, MN: Catalyst Press.

Pfannkuch, M., Budgett, S., & Thomas, M. O. J. (2014). Constructing statistical concepts through bootstrap simulations: A case study. In U. Sproesser, S. Wessolowski, & C. Worn (Eds.), *Daten, Zufall und der Rest der Welt—Didaktische Perspektiven zur anwendungsbezogenen Mathematik* (pp.191–203). Berlin: Springer-Verlag. doi: 10.1007/978-3-658-04669-9

Pfannkuch, M., Forbes, S., Harraway, J., Budgett, S., & Wild, C. (2013). *Bootstrapping students' understanding of statistical inference*. Retrieved from http://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report.pdf

Pfannkuch, M., Regan, M., Wild, C. J., Budgett, S., Forbes, S., Harraway, J., et al. (2011). Inference and the introductory statistics course. *International Journal of Mathematical Education in Science and Technology, 42*(7), 903–913.

Pfannkuch, M., Regan, M., Wild, C. J., & Horton, N. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education*, *18*(1), 1–38. Retrieved from http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf [Inaugural American Statistical Association Best Paper Award]

Pfannkuch, M., & Wild, C. J. (2012). Laying foundations for statistical inference. *Proceedings of the12th International Congress on Mathematics Education, Regular Lectures 1–9, 8–15 July, Seoul, Korea* (pp. 317–329). Seoul, Korea: ICME-12. Retrieved from http://icme12.org/

Pfannkuch, M., & Wild, C. J. (2014). Working together to improve statistics education: A research collaboration case study. *Proceedings of the 59th International Statistical Institute World Statistical Congress, 25–30 August 2013, Hong Kong, China* (pp. 608–613). The Hague, The Netherlands: International Statistical Institute. Retrieved from http://2013.isiproceedings.org/

Pfannkuch, M., Wild, C. J., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM—The International Journal of Mathematics Education*, *44*(7), 899–911. doi 10.1007/s11858-012-0446-6

Pfannkuch, M., Wild, C. J., & Regan, M. (2013). *Students' difficulties in practicing computer-supported statistical inference: Some hypothetical generalizations from a study*. In T. Wassong, D. Frischemeier, P. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen* [Using tools for learning mathematics and statistics], (pp.393–403). Wiesbaden, Germany: Springer Spektrum. doi: 10.1007/978-3-658-03104-6

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N.J. (2010). Inferential reasoning: Learning to 'make a call' in theory. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society: Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8), July 2010, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications.php

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference [with discussion]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(2), 247–295. [Paper read at the Royal Statistical Society in London, October 2010]

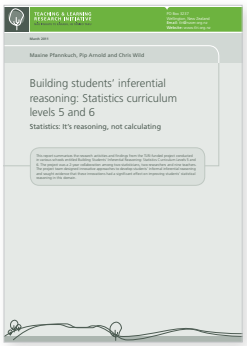
Wild, C. J., Pfannkuch, M., Regan, M. & Parsonage, R. (accepted). Next steps in accessible conceptions of statistical inference: Pulling ourselves up by the bootstraps. *International Statistical Review*.

An introduction to the selected projects

The details provided here serve as the briefest of introductions to the two highlighted projects in order to provide a context for the discussion that follows. Reports for both projects can be found on the TLRI website, and additional details can be gleaned from the extensive publication list (Box 1).

Building students' inferential reasoning: Statistics curriculum levels 5 and 6
Statistics: It's reasoning, not calculating

Maxine Pfannkuch, Pip Arnold and Chris Wild

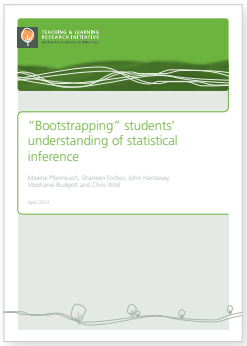


This project was a 2-year collaboration among two statisticians, two researchers and nine teachers. The project team designed innovative approaches to develop Year 10 and 11 students' informal inferential reasoning. These approaches involved building students' concepts and reasoning about samples, populations, sample and population distributions, sample size effect, and sampling variability in order to make informal inferences about populations from samples. The team sought evidence that these innovations had a significant effect on improving students' statistical reasoning in the domain of informal statistical inference.

http://www.tlri.org.nz/sites/default/files/projects/9275-summaryreport_0.pdf

'Bootstrapping' students' understanding of statistical inference

Maxine Pfannkuch, Sharleen Forbes, John Harraway, Stephanie Budgett and Chris Wild



This project was a 2-year collaboration among three statisticians, two researchers, sixteen Year 13 teachers, seven university lecturers, one workplace practitioner, three teacher professional development facilitators, and one quality assurance advisor. Using non-traditional bootstrap and randomisation methods the project team designed innovative, computer-based approaches to develop Year 13 and first-year university students' inferential reasoning. The team sought evidence that these innovations were effective in developing students' understanding of statistical inference.

http://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report_0.pdf

The main players

The partnership principle of the TLRI was embedded in both projects from the leadership level down. Both projects were co-led by Maxine Pfannkuch, who is a mathematics teacher educator, and Chris Wild, who is a research statistician and teacher of tertiary statistics courses. Both Maxine and Chris work at the University of Auckland. Their long-standing association began when Chris was one of Maxine's PhD supervisors. Since that time they have continued to share a deep interest in how best to teach so that students develop strong capabilities in statistical reasoning. From their different backgrounds (Maxine was a secondary school teacher in her first career), each of them brings a different perspective to their research on statistical reasoning.

Other partners brought further perspectives to the work in both projects. Some participants were advisers who worked with mathematics teachers in schools (e.g. Pip Arnold), and some were professional statisticians who brought employment perspectives. As the above summaries show, partnerships with classroom teachers were also an important feature of both projects. The power of these multiple contributions and perspectives will be evident throughout the discussion that follows.

Keeping track of the action

Three interrelated timelines can be found on page 38 for quick reference. The first outlines the timing of relevant research activities. The second details national networking activities of the TLRI team on each project. The third outlines key policy-related activities. A common time scale aligns the three pages so that it is possible to see at a glance how the different types of activities co-evolved.

Doing research that matters

Researchers do, and should, choose topics that interest them personally. However, it is also important to be clear about *why* new knowledge in the focus area is needed. What difference will it make to practice? Why and for whom will that difference matter? What makes the investment of researchers' efforts—and TLRI funds—worth making? I think of this as the 'So what?' question.

Compelling conception of ends: *Those involved are clear about how and why addressing the challenges of a specific project could help to generate knowledge that advances practice. Both depth and breadth of relevant expertise are needed to successfully address the ends envisaged.*

It is noteworthy that Singer-Gabella presented this as her first criterion for demonstrating scholarship in practice. Scholars should have a *compelling conception of the ends* they are working towards. They bring the breadth and depth of their knowledge to bear on a challenge about which they care deeply, and think others should too.

In their own words....

The quotes that follow have been drawn from the statistical reasoning teams' written accounts of their work. Some are slightly abridged (e.g. citations to other research papers have been left out, but can be found in the originals). I made this selection to illustrate how and why the two TLRI teams saw their work as important and urgent.

> There was an urgent need to understand how students can progressively develop informal statistical inferential reasoning from Levels 5 to 8 of the new curriculum. (Quote from the TLRI website page for the first project)
>
> The gap between statistical practice and statistics education is increasingly widening. The use of new computer based statistical inference methods using re-sampling approaches is pervading practice. However, statistics education remains trapped by what was computable in the 20th century. (Quote from the second TLRI project report, p. 2)
>
> In most introductory statistics and Year 13 courses the conceptual foundations underpinning inference are the normal distribution, the Central Limit Theorem and the sampling distribution of estimates. Research evidence, however, suggests that these theoretical and mathematical procedures act as a barrier to students' understanding, and the statistical inference concepts are inaccessible to the majority of students. (Quote from the second TLRI project report, p. 2)
>
> In a society that demands evidence-based thinking, and therefore that its citizens become statistically literate, there is now an urgent need for more research into growing young people's statistical reasoning. (Concluding statement from the first report)

The researchers note the inclusion of inferential reasoning in the Mathematics and Statistics learning area of *The New Zealand Curriculum* (*NZC*, Ministry of Education, 2007). But their concerns are not driven by compliance; i.e. "We need to do it because it's in the curriculum". Indeed, as the following sections of this report will show, key members of both project teams were actively involved in influencing the development of the achievement objectives in the curriculum.

A compelling democratic argument is implied in the selected quotes. The researchers see an urgent need to ensure that as many young people as possible become statistically literate in a society that places a premium on evidence-based thinking. In a world of 'big data', students need to learn to use and understand the computer-driven methods of contemporary working statistics. One obvious teaching implication is that we need to find ways to help a much wider range of students—not just the most mathematically able—to learn to make sound statistical inferences in many different contexts. But current

evidence suggests that most students are unsuccessful in achieving these outcomes by traditional computational means. This implies a need to find new ways to teach students how to make these kinds of inferences. Any new approaches need to be built on a robust knowledge of how students' ability to make inferences develops over time, and how this development can be supported by strategically chosen learning experiences.

Notice how concerns about teaching *and* learning are tightly interwoven in this summary of the driving rationale for both projects. The challenge posed cannot be addressed by finding out more about teaching *or* more about learning. New knowledge is needed in both these areas. This type of complex inter-relationship between teaching and learning is a hallmark of scholarship in practice. It is also the kind of knowledge that TLRI projects should be seeking to generate, because it is the sort of knowledge with the strongest potential to make a real impact on practice.

Notice the concern for students *in their futures*. The learning envisaged is not just for today (e.g. for making short-term gains and passing assessments). Instead, the learning matters in some bigger sense for students' lives beyond school. This dual concern for both the present and future salience of learning has also been noted as a hallmark of the pedagogy of teachers who provide strong support for students to develop the *NZC* key competencies (Hipkins, Bolstad, Boyd, & McDowall, 2014).

> It wasn't just another research project. Everyone in the team had a strong commitment to learners in New Zealand. (Teacher participant)
>
> I've always been committed to the compulsory sector: making a difference for kids in New Zealand, keeping kids doing maths for as long as possible at school. (Adviser)
>
> I was just starting to think about how important statistical literacy was as a life-skill when the opportunity came along to get involved in the first project. (Teacher participant)
>
> It's widely accepted that statistical literacy is crucial for being an informed participant in society and in higher education. (Adviser participant)

This sense of fostering cogent, memorable statistical reasoning strategies which students will take forward into their futures is evident in the visual image in Figure 1 on the next page. This image comes from the first of the TLRI reports and emphasises that different samples from the same population will show differences in their distribution patterns and that these differences must be taken into account when making an inference.

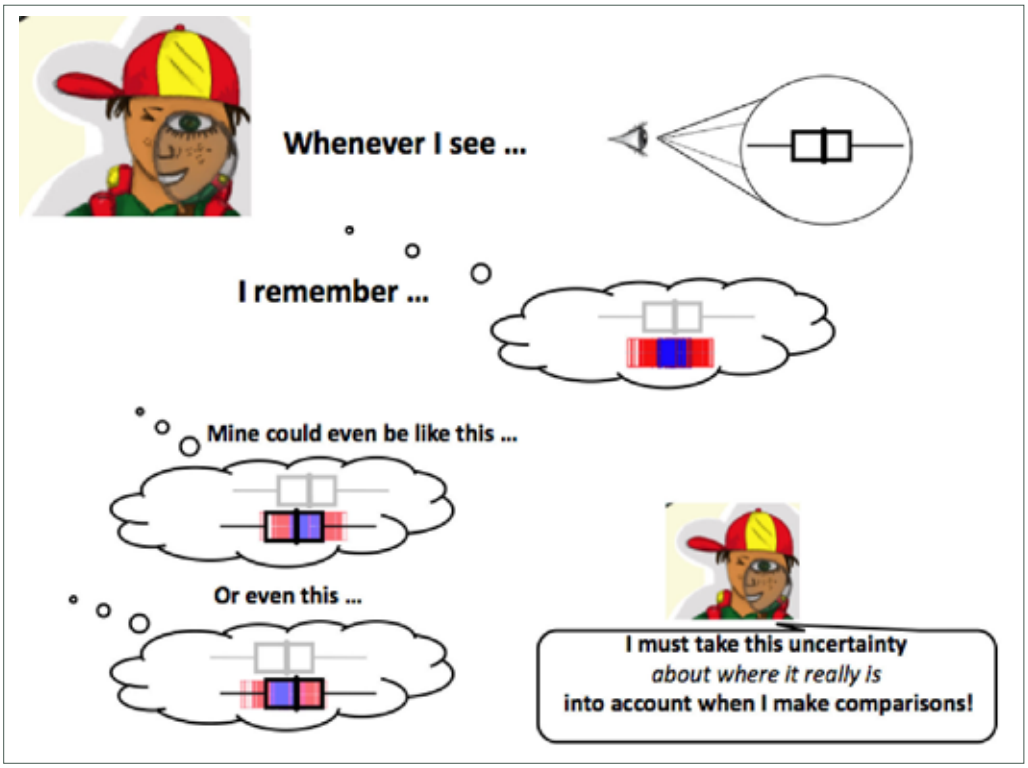


Figure 1: **What the researchers want all students to take forward into their futures**

Keeping a tight focus for manageability and impact

Making inferences is just one dimension of statistical thinking. I asked the project leaders why they picked this specific focus for two consecutive projects.[2] Their responses made it clear that this was a carefully reasoned choice. They argued that students' success in statistical inquiry can be more strongly affected when the focus is on inference compared with other aspects of the statistical enquiry cycle. Making appropriate inferences is obviously important when data have been gathered and are being interpreted. However, the project leaders reasoned that a robust understanding of inference is also a key to success at the design stage. As Chris noted, "once the data are in it's too late to go back."

Good design of a statistical investigation always begins with an acute awareness of the context of the planned inquiry. Figure 2 shows that the research team had a very clear view of the relationships between learning to think critically about the *inquiry context* and making defensible inferences. Notice how clearly this diagram conveys relationships between all the aspects of reasoning and thinking the researchers wanted to foster.

2 In a third TLRI project, currently underway, the team has moved on to investigate the teaching and learning of probability.

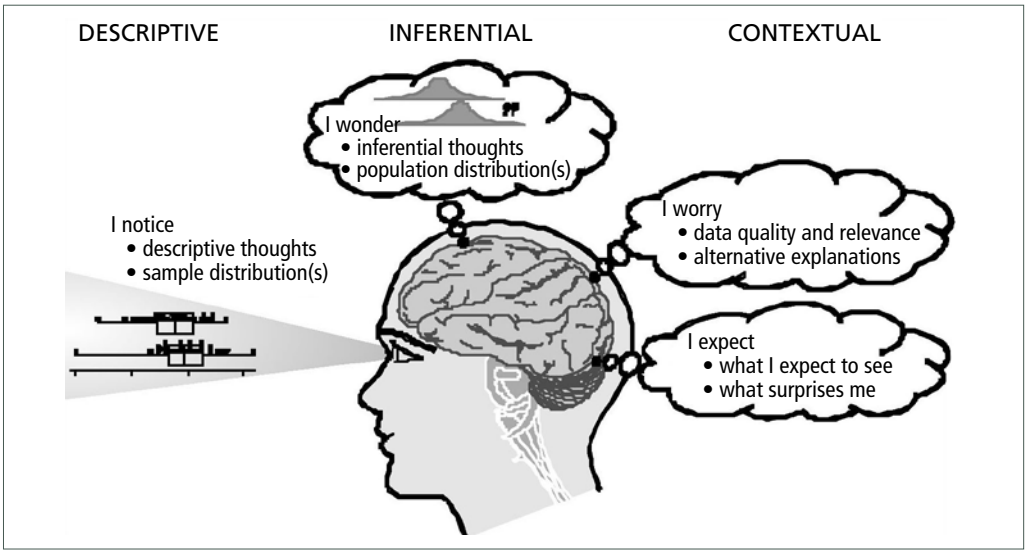


Figure 2: **How the researchers modelled desired reasoning and thinking**

Strong interest from the international research community also supported the chosen focus and helped the research team build a compelling case for the projects. In 2005 Maxine presented research generated from her PhD thesis at the biennial International Statistical Forum. Her work on the nature of statistical thinking had led her to question how much was actually known about whether and how students differentiate between observation and inference. The case she made was obviously compelling: making inferences became the main focus of the 2007 forum. Within the collective of international members represented at the forum, recognition was building about the need to know more about the conceptual steps that students take when making inferences, and what conceptual development looks like over time.

Expanding the focus to include bootstrapping and the randomisation test

Bootstrapping is widely used by professional statisticians. It addresses the problem that there will always be sampling error in estimates produced by drawing random samples from populations. While the extent of the sampling error due to chance is initially unknown, the computer-driven technique of bootstrapping allows this unknown to be systematically investigated for sample-to-population inferences.

The randomisation test addresses a related but different question. Can variation between groups be attributed to chance alone or is something else going on? This question is central to any statistical study that involves a comparison between treatment and control groups, so it is important for students to understand the challenge of experiment-to-causation inference.

I was aware of various traditional randomisation techniques but I had never heard of bootstrapping before I began to write about this programme of TLRI research. I expect many readers of this report will be in the same position. So why did this conceptually challenging idea become the focus for the team's second TLRI project? As I now briefly outline, the focus of the second project emerged from the insights generated from the first.

Drawing repeated samples from the same population introduced the basic concept of sampling error in the first TLRI project. This was done initially in a very hands-on way, with students drawing individual cards from a bag of cards that represented the whole population and comparing the samples they arrived at. This early, very concrete work progressed to the types of computer-generated visuals and judgement calls illustrated in Figure 1. This first project had shown that students with a wide range of mathematical abilities could successfully learn how to 'make a call' based on *visual displays* of sampling variability.

The opportunity the team now saw was that bootstrapping produces visual displays that look very similar to those created by repeatedly sampling the original population. But there is a catch. Bootstrapping actually *re-samples a sample* from the population. The process parallels the resampling from the whole population, and the visual display looks very similar. The team initially worried that this would be too confusing for students, but concluded that "the similarity between these patterns is why bootstrapping makes sense and why it works" (Wild, Pfannkuch, Regan, & Parsonage, 2013, p. 13).

By teaching bootstrapping, the team reasoned that they would be giving students vivid and memorable insights into the actual processes used by research statisticians to check sampling error, and that are increasingly being included in statistical analysis packages for lay use. Thus the motivation to focus on bootstrapping in the second project again related to the team's compelling conception of developing statistical literacy that would abide well beyond school learning.

> Bootstrapping was developed in the 1970s. It is brilliant common-sense. The rapid computer development of the 1990s enabled accessibility to this idea outside the professional statistics community. (Research statistician)
>
> The processes used to randomise data should also be the processes used to analyse sampling variability. Bootstrapping does this. (Project leader)
>
> In the first project we didn't know that our next focus would be on bootstrapping. From my earlier research we knew the idea of resampling was important: it all grew from those marks on the whiteboard [see page 17]. (Project leader)
>
> Learning about resampling is an objective in *NZC*. When people were working on the new achievement standards in the alignment process, the New Zealand Statistics Association (NZSA) education committee, of which we are members, was giving feedback to the writers, commenting on the areas being developed. This is how bootstrapping got in there as an internally assessed standard at level 3 for that objective. (Project leader)
>
> In the second project team, one teacher had studied bootstrapping and the rest of us didn't know about it. We were learners ourselves at first. (Teacher participant)

This sequence of comments demonstrates once again the team's commitment to support learners to broadly understand the conceptual underpinnings of statistical inquiry as it is actually practised in real world settings. When this vision logically pointed towards a process that had never before been part of the school curriculum, they had the courage to lobby for its inclusion. They then followed through with the commitment to do the research needed to support teachers to bring these new and unfamiliar ideas to life in the classroom. I'll return to the relationship between research and policy (specifically curriculum and assessment policy) shortly.

Designing convincing investigations of practice

The second criterion for scholarship in practice is that appropriate means or methods are used to achieve the compelling ends envisaged. These means might include the design and refinement of new resources intended to support the achievement of the project's goals. The design process is often recursive as evidence emerges about whether and how any such resources support the project's ends. All these features can be seen in the TLRI statistical reasoning projects.

Appropriate means or methods: *The means or methods employed are appropriate to the intended ends and are deployed in an "intentional, skilled, dynamic" manner. The process used is often recursive, as evidence emerges about whether and how the means support the project's ends.*

Choosing an appropriate methodology

Making an appropriate choice of methodology is critical to answering a project's research questions. This probably seems so obvious that it hardly needs saying, but experience suggests that it can be easier said than done. In the case of the two statistical reasoning projects, a *design research* methodology was proposed. Here is how the team explained this choice:

> Design research aims to develop theories about learning and instructional design as well as to improve learning and provide practitioners with accessible results and learning materials. (Quote from the second TLRI report, p. 3)

Notice that this explanation encompasses both teaching and learning. The design challenge here is to figure out potential obstacles to learning progress, and to design and test ways to overcome these by using effective teaching approaches. This aim can only be fully met by gathering robust evidence of learning challenges, and evidence of learning gains that show these challenges have been overcome. This must be achieved in ways that allow these gains to be correlated with the teaching and learning approaches and materials being investigated.

Directing design efforts where they will have most impact

The in-principle choice of design for the research is just the first step. The next step is considering what needs to be (re)designed and why. Again this is a deceptively simple question. The project leaders' explanations for their choices illustrate how deep knowledge of existing practice can inform design thinking about what to change and why.

Before the TLRI projects were designed, Maxine had spent time working in one or two classrooms, experimenting alongside the teachers as they made some changes to the ways they taught statistical reasoning. One day, in a spontaneous response to the learning challenges she was picking up, Maxine drew a series of marks on the whiteboard in rapid succession. These marks represented variation in the medians of different samples drawn from the same population. The effectiveness of this tactic alerted her to the role that visual displays might play in helping students understand sampling variability.

Chris's experience provided a complementary dimension to the focus on the impact of visual displays. As we saw in the first section of this report, the team was very aware that computers had transformed the investigative processes used by working statisticians. They also knew that new IT tools were "beginning to place powerful, accessible, visually-based data-assembling and analysis tools right into the hands of primary decision-makers themselves" (Wild et al., 2013, p. 1). This implies that a much wider group of people now need to know how these tools work.

The extent and power of these changes convinced the project leaders that students should also have opportunities to use enquiry processes that mirror those undertaken by statisticians and, increasingly, by other adult decision makers. Chris had recently read a book on visual perception that reinforced his belief in the importance of visual displays and visual meaning making.

Once the idea had formed, the team searched for recent research literature that might explain how/why visual displays make a strong impact on learning. This literature, in turn, led them to recent insights into how the brain functions in cognition. And this, in turn, led to the development of a set of principles to guide the design of their own learning materials and approaches. Broadly paraphrased,[3] some of the design principles addressed important matters such as:

- **avoiding cognitive overload**—most of the visual display remains as expected while one key element changes during the animation
- **directing attention to salient features**—colour and bolding are used to highlight important parts of the visual display, while movement is used to draw the eye, and fading can be used to show when to stop paying attention to a specific element and refocus elsewhere
- **building familiarity before introducing new concepts**—simple hands-on activities, user controls for the animations and self-pacing controls allow students to build familiarity and basic comprehension before new concepts are represented
- **combining pictorial, verbal and movement elements in key actions**—for example, 'learning to make the call' was the phrase used to cue a focus on the need to think carefully about the meaning of variability displayed by two samples;[4] making the call initially involved simple hand actions, which proved to be very effective in consolidating the intended learning, and supporting subsequent recall of the critical messages.

Chris used his own programming ability, combined with his knowledge of computer-based enquiry techniques, to write simple programmes that teachers could trial in their classrooms. Across several design iterations, these programmes provided strong visual displays of sampling variability. Students could observe the effects of multiple repeat sampling for a population simply by attending to how the display changed from sample to sample. (Figure 1 gives an indication of how this works.) And, as already noted, once the prototype displays had been created and used effectively, the connection to bootstrapping and the randomisation test became the clear next focus for the second project.

3 The full set of principles is outlined and explained in Wild et al., 2013.

4 The phrase was adapted from the comment made by a student in one of the teacher participant's classes that a specific judgement was "too close to call".

Gathering clear evidence of the impacts on learning

It was important to the team that they could clearly document the impact of the teaching and learning strategies they had devised and supported with new innovative resources. Could they find evidence of strong learning gains? In order to answer this question they needed to develop clear criteria for what these learning gains might look like. For the first project they did this by developing detailed assessment frameworks for five domains of reasoning they saw as important components of making inferences. Notice how every component in Figure 2 is represented in the set of criteria.

- **Making the call:** Are samples from two populations sufficiently different that we can infer an actual difference in the sampled populations? The 'call' in question here requires students to differentiate meaningful differences from random, unavoidable sampling variation. It is represented in Figure 2 by the 'I wonder' bubble.
- **Shape:** Can students generalise the overall shape of the distribution of data in a sample set? This is represented in Figure 2 by 'descriptive thoughts' under the 'I notice' heading. This question came more clearly into focus in the second year of the first project, when the team realised that students did not have a good sense of how to create and interpret generalised distribution shapes. This design addition illustrates the value of conducting iterative cycles of design research.
- **Spread:** The criteria for this domain of reasoning address the development of students' ability to describe the overall distribution of a sample. Again this is represented in Figure 2 under the 'I notice' heading. Spread has traditionally provided the computational avenue for developing awareness of sampling variability (e.g. mean, median, mode, standard deviation, box-and-whisker plots, etc.). In the vision of the TLRI team, it is initially represented visually so that the concept is understood before computation is introduced.
- **Context variability:** Teaching students to differentiate between the variation that occurs naturally in populations and the variation that arises from sampling those populations was central to the first project. Awareness of the types of variability that might be anticipated, given prior knowledge of the context, is represented in Figure 2 by the 'I expect' speech bubble.
- **Unusual patterns:** Critical thinking about the context in which the data were gathered, and the process used to sample the population, is made explicit in this set of criteria. Figure 2 represents this focus in the 'I worry' speech bubble.

Two sources of evidence were used to make an overall judgement about where individual students were positioned with respect to these criteria before and after the teaching sequences. All students completed pre- and post-tests, and some were interviewed about their responses to these pre- and post-tests. Several members of the research team, including some research assistants, independently rated students' test responses against the criteria. Inter-rater comparisons were then used to ensure the team had a shared understanding of what the criteria meant. Interview responses informed this discussion by clarifying the type of thinking that underpinned specific types of responses. This robust combination of analytical activities allowed the team to make confident claims of learning gains at the end of the first project.

A similar process was followed in the second project. This time the specific criteria developed related to the conceptual challenges of interpreting real-life studies to test specific knowledge claims (e.g. using treatment and control groups in medical investigations).

Resources as products of scholarship in practice

TLRI projects are research projects. First and foremost they are designed to build new knowledge about teaching and learning. If a proposed project wanted to create new resources for their own sake, it would not be successful in winning funding. But *design research* often requires resources that don't yet exist. Before the teaching and learning ideas can be tested, these resources will need to be developed. They are a legitimate part of the research process in these cases.

Sound judgement:
Scholarship in practice demands that both theoretical and practical knowledge are brought to bear, carefully grounded in the contextual specifics of the challenge being addressed. Choices and their consequences are made visible so that the products of the scholarship can be more readily transferred to other contexts.

Creating resources to test new teaching approaches

As we have already seen, one working hypothesis for the first statistics project was that visual displays would enable students to develop conceptual understanding that might not have been accessible via more traditional computational routes. (In other words, their maths wasn't strong enough to develop the ideas via traditional maths activities.) This hypothesis could not be tested until the visualisations had been created. Once Chris had completed the relevant programming, patterns in sampling variation could unfold before students' eyes. Criteria for learning gains also had to be designed: only then could the hypothesis be empirically tested.

The visual resources generated during the first project were extended to include simulations of bootstrapping in the second project. Again, it was an open question whether students would be able to develop the types of statistical reasoning the resources targeted. Exploring the *learning consequences* of students' interactions with the resources was central to the investigation. So, again, developing the resources was a necessary step in an integrated sequence of research activities.

Resources as 'things to think with'

As paraphrased above, the third criterion for scholarship in practice includes the idea of *transfer* of new knowledge generated by scholarly activities. Products of scholarship in practice have an important role here when they are used to spread new practices beyond the original team. Attractive and compelling new resources (i.e. such that teachers want to use them once they know how) are obvious candidates for the spreading of new ideas. With this transfer imperative in mind, it is important to note that the visualisation programmes Chris created are now available for any teacher to use.[5]

New resources can act as *thinking objects* when they confront teachers with unfamiliar ideas or surface tacit thinking, and/or they challenge current beliefs and practices. Critical reflection about current thinking and practices is known to be a hallmark of effective professional learning (Timperley, Wilson, Barrar, & Fung, 2007). In the case of the visualisations, some teachers were challenged by the claim that students can learn important statistical concepts without needing to use traditional computational methods. But as the title of the first report emphasised, "It's reasoning, not calculating".

5 They can be accessed via the CensusAtSchool website: http://new.censusatschool.org.nz/

The teachers who worked on the project certainly understood this. They were also aware that using the resources could be the compelling first point of contact with this challenging idea for teachers who would not necessarily engage with research.

> Resources contribute to teacher buy-in and success. For some teachers, using the resource opens up their thinking, then they engage with the research later on. (Teacher participant)
>
> Different kids get to shine when they use the data to tell stories. 'I notice, I wonder' fits lots of maths contexts. (Teacher participant)
>
> Thanks to the brilliance of Maxine and Chris, the technology allows kids to do what real statisticians do. (Teacher participant)
>
> The Year 9 and 10 resources show how to scaffold conversations. You need to learn how to manage cognitive dissonance: to listen and facilitate students' developing understanding. It's not what maths teachers are used to doing. (Adviser participant)

Ensuring the products of scholarship are used appropriately

Teachers who participated in the projects saw at first hand how and why the supporting resources were created. Their personal professional knowledge and judgment made an important contribution to the practicality and usability of the classroom materials. But this was a small group relative to all the teachers of statistics in New Zealand. It was important to the team that other teachers also understood the nature of the learning the resources supported and how to use them appropriately to achieve the anticipated outcomes.

The team began the process of reaching out to other teachers by using their networks. There was an initial focus on using existing professional learning networks in the Auckland area, such as the Saturday morning workshops, held once a term, organised by the Auckland Mathematics Association (AMA), and the annual Statistics Teacher Day hosted by The University of Auckland in conjunction with the AMA. The TLRI team made effective use of these opportunities to introduce their work to a much wider network of teachers. Comments made by different members of the wider team were an indication of the importance attached to this first-hand dissemination and its role in ensuring the resources were used as intended (both project leaders have been involved in giving Saturday morning and Statistics Teacher Day talks).

> For sustainability, you have to be involved with the community of teachers. (Project leader)
>
> A strength of the TLRI is that it reaches teachers on the periphery, not just the teacher researchers. (Adviser)
>
> It's not just about 'you're a good teacher, you can do this'. You've got to know why and where you are going. (Adviser participant)
>
> The Saturday morning workshops have become a key part of getting the message out. (Adviser participant)
>
> The vocabulary we learned to use empowered me to explain the results of a randomisation test. (Adviser, who was a teacher at the time)
>
> If you can make the change before they [teachers] have done something wrong, you've got there. (Adviser participant)

Transferring these ideas to other projects

In summary, all of the following can make an important contribution to the demonstration of scholarship in practice:

- creating resources to address specific learning challenges
- testing and refining these resources in practice
- using new resources as aids to the wider dissemination of challenging ideas and new pedagogies.

All of these activities demand a skilful weaving of theory and practice. However, wider uptake of new approaches and resources will not happen unless the resources make sense to teachers, the conditions of their use seem achievable and the effort of making the change seems worthwhile.

Box 2 distils this summary into four questions that might help other research teams in deciding whether resource production is an appropriate part of a proposed research project.

Box 2: Generalising the resource production question to other projects

- Will the resource illustrate the scope and reach of a new idea? How will it do this?
- Will the resource open a space for creative imagining of new possibilities? (Who will be involved in this imagining and how will they do it?)
- Will the resource act as a 'thinking object' for teachers or other partners? (Could use of the resource reveal specific instances of cognitive dissonance to prompt a rethinking of personal ideas/practices?) Who will scaffold and prompt the thinking needed to reveal any dissonance in beliefs and practices?
- Could the use of the resource allow practical barriers and challenges to emerge and be debated? Who should participate in these practical learning conversations?

Advancing the knowledge and skills of teachers

Significant outcomes: *Scholarship in practice produces local and immediate outcomes that advance the knowledge and skills of teachers and other educators who are directly affected by the work.*

The fourth criterion for scholarship in practice is that there are *significant immediate outcomes*: the knowledge and skills of teachers are enhanced in ways that lead to changes in their practice. This means the research needs to be carefully grounded in the day-to-day realities of current practice, even as it seeks to transform teaching and learning. Aspects of this challenge have already been addressed in the discussion of the part played by resources. Now I turn to one particularly salient TLRI principle:

> Principle 4: The research projects within the TLRI will be undertaken as a partnership between researchers and practitioners.

The TLRI requirement that projects be designed as genuine partnerships enhances the likelihood of success in meeting the above criterion for scholarship in practice. However, if teachers are positioned as research *subjects* rather than genuine research partners, it is most unlikely that transformation of practice will actually happen. (A typical scenario here would be that the researchers direct the teaching action and then evaluate how well the teachers do what the researchers want.)

When *partnerships* between researchers and practitioners work well, new knowledge can be created 'in the spaces between' the different expertise the partners bring. Maxine's small-scale projects, beginning in the 1990s (see the timelines page 38–39) provided an important platform for building respectful and productive partnerships with secondary teachers. They gave her a feel for good ways to get teachers actively engaged and wanting to find out more. She noted the power that came when teachers saw their students do things they had not thought would be possible. Success fuels success!

The projects from the participating teachers' perspectives

I interviewed three teachers who took part in one or both statistical reasoning projects. As the following quotes show, they saw themselves as active partners in the research and were very clear about their specific place in, and value to, the overall programme of work.

> I liked being valued for my work. Making connections was really good.
>
> I went in with 'old ideas' and was challenged, but I never felt I couldn't argue my case.
>
> Make sure someone on the leading team understands high schools. They are quite different environments from universities: different pressures; different pedagogies; different class sizes.
>
> I always felt I could say 'that's not going to work and here's why'. Even if no change was made in the end, you always knew why.
>
> We have to teach things we never learned [bootstrapping]. Not everyone can do that.
>
> We weren't just some teachers. We were talented statistics teachers with strong guidance from the statisticians in the team.
>
> We have to make better statisticians of our statistics teachers.
>
> Make it valuable for us and teachers will give, and give, and give.

These quotes show that the secondary teachers were clear that they needed to—and were expected to—bring the grounded realities of the classroom to the research. They

held practical knowledge that others could not contribute as cogently to the team. However, they were also clear about the learning they needed to undertake, and how they could benefit from the different expertise the university researchers brought to the table.

Supporting the professional learning of their peers

All the teachers took the new ideas back to their own school teams. The teachers who did not live in Auckland had the additional responsibility of taking new ideas, methods and resources back to all the other teachers in their regions. It really helped that they were strongly networked into their local teacher associations and were confident to organise and run professional learning for their peers. One teacher said that having advisers in the team was helpful here. From their own experiences, the advisers willingly shared ideas about the best ways to engage other teachers as learners. One participant teacher had copied the sequence of learning activities used in the Auckland meetings, replicating these in the local region. Regional teacher networks repaid this support in kind when they helped with travel and accommodation costs for the project meetings in Auckland.

Leveraging other professional learning opportunities

The project leaders' commitment to teachers' ongoing professional learning extends well beyond the scope of their own immediate research interests. They have shoulder-tapped and encouraged teachers to apply for other professional learning opportunities, such as the Endeavour Teacher Fellowships managed by the Royal Society.[6] At the time of preparing this report, one of the teacher participants from the second project was undertaking such a fellowship. Her project aimed to identify rich data sources that any secondary schools could use for statistical investigations. Another teacher participant had just begun to study for an MEd at a university in his region.

Other benefits for teacher partners

Being involved in research, with all its uncertainties and challenges, has given the teacher–partners greater confidence to deal with uncertainty when it arises in their own classrooms. Such confidence is especially pertinent in real-world statistical inquiries. One teacher laughingly recalled a recent conversation in which a Year 13 student noted that the data they had been given looked "really ugly". This observation led to a rich conversation about what to do with conflicting data. The teacher felt that a lot of other teachers would struggle with such an open-ended conversation where there was no clear right answer.

One of the teachers also discussed the value of having several years' advance warning to think about the wider implications of the curriculum and assessment changes that were coming. First-hand experience of using the software resources had prompted him to lobby for a whole new computer laboratory just for the statistics classes in his large secondary school. He noted that the students really enjoyed the new approaches and working in the computer lab, which was now full most of the time.

6 http://www.royalsociety.org.nz/teaching-learning/teacher-fellowships/endeavour-teacher-fellowships/

Conceptualising relationships between partners

The ongoing programme of TLRI projects affords an opportunity to interrogate the conditions that have an impact on the effectiveness of research–practice partnerships. One paper I found while preparing for this project (Penuel, Coburn, & Gallagher, 2013) discusses the important role played by people who can act as go-betweens. These are people who can walk in the worlds of at least two participant groups and hence keep knowledge circulating where and when needed. I asked the project leaders whether they thought this observation applied to their two completed TLRI projects, and if so, who the go-betweens were. Interestingly, they didn't see the relationships they fostered in these discrete terms: 'go-between' implies that there are separate groups that need to be deliberately connected. Instead, they thought of the interconnections between the different participant groups as constituting "overlapping spheres of influence".

This section has mainly focused on learning impacts for the teacher partners because their voices have not been heard as often in the earlier sections. It is important to note, however, that the criterion for scholarship in practice certainly does not imply that the learning is a one-way street. As the paraphrased quote that begins this section says, such scholarship advances the knowledge and skills of teachers *and other educators who are directly affected by the work.* Since everyone involved in the work is an educator in at least part of their work, this implies that the conditions should be such that everyone gains valuable new insights by being a participant. The metaphor of partnerships as overlapping spheres of influence succinctly captures this wider sense of reciprocal learning and shared responsibility for ensuring the success of the work.

Leveraging change at the policy level

This fifth criterion for scholarship in practice casts a very broad net. Research has a demonstrable impact when it opens up new avenues of enquiry. This is one familiar hallmark of traditional scholarship, and as we have seen, these two TLRI projects have been very successful in these terms.

However, the criterion also encompasses the influence of research on policy directions. This avenue of influence has not always been associated with the idea of scholarship in the traditional sense. With the idea of scholarship in practice to the fore, this section explores ways in which the TLRI researchers strategically leveraged policy opportunities to connect research and practice. The generative power of doing so was a clear theme of the interviews I held with different members of the project team.

Generativity: *This criterion shifts the focus to longer-term and less immediate outcomes. The work has impacts beyond the immediate context. It opens up new ideas and directions of enquiry. The work speaks to a wide range of stakeholders: teachers and teacher educators, educational administrators, and education policy makers.*

Leveraging curriculum change

> Maths teachers love the abstract—pure reasoning. Now students have to reason in 'fuzzy' situations, underpinned by a focus on sampling variability. This has never been a focus in the past and now it's at the heart of *NZC* objectives. (Adviser)

At around the time that Maxine was experimenting with marks on the whiteboard to help students understand sampling variability, she was also subcontracted to lead the development of the statistics strand of the mathematics learning area of *The New Zealand Curriculum* (*NZC*) (Ministry of Education, 2007). The process of updating curriculum policy and supporting materials often falls to teams of highly experienced teachers who bring a wealth of practical wisdom to the process. These teachers may not, however, be aware of new curriculum directions that could be about to open up. Researchers with a strong theoretical orientation are more likely to be aware of imminent challenges of this type. This was certainly so for the statistics team at Auckland University and among the wider group of working statisticians with whom they consulted. It was clear that the project leaders saw a strategic opportunity to lead curriculum thinking.

> We tried to be forward-looking and anticipate future needs. (Project leader)
>
> We knew how fast statistics itself was changing. (Project leader)
>
> The leading mathematics teachers knew, from the 1992 mathematics curriculum, that mathematics could be defined in the framework of mathematical processes—logic, problem-solving, communication—in contexts that matter. When the curriculum review of the 2000's got going, these leading teachers were ripe for conversion. They could see that statistical thinking is about similar but complementary processes. (Research statistician)

Under the auspices of the New Zealand Statistics Association Education Committee, Maxine pulled together a large team to help with the curriculum development work. Together they developed the conceptual ideas that were written into the achievement objectives for statistics. Because they were breaking new ground, these *NZC* achievement objectives described learning challenges that were unfamiliar. Teachers wouldn't be able to draw on their existing pedagogical content knowledge when working out what to do, how to sequence learning, how to recognise and overcome potential learning obstacles, and so on. The team were very aware of the need for advice and support, but they didn't yet know themselves how best to support students' learning so that the objectives would actually be achieved.

> We had a problem! No-one knew how to do it. Could kids actually do it [what was now specified in the NZC learning objectives]? This was an open question at the time. (Project leader)

This dilemma provided the genesis of the questions for the first of the two TLRI statistical reasoning projects. In addition to the research-facing imperative, Maxine and Chris were very aware that most teachers would not adopt the new achievement objectives in the absence of criteria that showed them what students were actually expected to do. The practice-facing imperative for conducting the research was equally strong and provided the impetus for a focus on matters such as what to look for when assessing evidence of progress.

Leveraging assessment change

> Have a finger on the pulse of curriculum. Capitalise on curriculum and assessment change to achieve something positive for students and teachers. (Advice from one of the participating teachers)
>
> In the first project we were just winding up. We really got into it in the second one. We targeted the work to be up and running the year before it was needed [for new assessments]. (Comment from one of the advisers)

The alignment of assessment practices with new curriculum directions is an important enabler of curriculum change. In particular, changes to achievement standards[7] in the senior secondary school are often perceived to be high stakes for both students and teachers. For this reason, keeping both curriculum and assessment policy in play is one way of ensuring that teachers take up new ideas 'naturally' as they engage with the personal professional learning necessitated by assessment changes.

Again the leaders of the TLRI statistical reasoning team were very clear about the need to strategically leverage assessment processes to consolidate the focus on statistical reasoning. Pip and other members of the research team were involved in the standards alignment project that followed the development of *NZC*. The alignment process required all the achievement standards for NCEA to be revised, updated or, if necessary, rewritten. Since statistical inference was such a new area in *NZC* itself, totally new achievement standards needed to be developed, as did all the supporting resources. The team approached this challenge strategically, with the realities of teachers' working lives firmly in mind. They began the alignment project by working with the achievement standards that needed the least change so that they could build confidence before they moved on to the big new ideas.

Differentiating achievement within a standard

Developing criteria for a specific achievement standard is a demanding design task that requires deep knowledge of the likely trajectories from an initial grasp of an aspect of learning to deep and connected understanding. Such trajectories allow the standards writers to clearly differentiate between criteria that describe achievement, achievement with merit and achievement with excellence. The statistics team used the SOLO[8] taxonomy to support the initial design work. When it subsequently became clear that

7 Achievement standards are used to judge achievement and gain credits towards a National Certificate of Educational Achievement (NCEA), which is New Zealand's senior secondary school qualification

8 Structured Observations of Learning Outcomes (Biggs & Collis, 1982).

some of the achievement standards needed refining, emergent knowledge from the TLRI research was available to underpin and justify the changes made.

Working with the exemplar writers

Once the actual achievement standards had been developed by the team, the New Zealand Qualifications Authority (NZQA) commissioned other teachers to write exemplar tasks. One of the advisers noted that these exemplars "breathe life into the bare bones of the curriculum and standards". This makes them critically important for transferring the curriculum intent into actual classroom practice.

Maxine noted that it took a lot of effort to find a way to work with the writers because NZQA policy is not to share the names of people who do contract work for them. The project leaders persisted because they saw it as important to work with all the teachers who were shaping exemplars for internally assessed standards. They also wanted to ensure that examination writers were well briefed on the scale and the essential nature of the changes as they developed externally assessed tasks and schedules for making judgements about students' work. In this way, the TLRI team's commitment to working in partnership extended well beyond the initial research teams.

Developing senior secondary subject guides

Completing this thread of strong practical support to ensure curriculum and assessment policy were well understood and applied, Pip led a Ministry of Education contract to develop an online subject guide for teaching statistics at the senior secondary level.[9] Here, again, the influence and impact of the TLRI work spread far beyond the initial project team. Again, taking on this additional work afforded opportunities for the team to make adjustments in support and advice as the new policy directions bedded down and classroom practice began to change.

Marshalling the support of powerful allies

> Unlike many fields, teaching is one experience that everyone has some knowledge of—even if only from having been on the receiving end. Members of the community, teachers as well as politicians and educational administrators all have a view of what education is and how it does or 'should' work. As the primary reference for what a school 'is' or 'should be' is based on most (non-student) stakeholders' personal experience, this can act as a significant conservative force to the extent that those stakeholders find it difficult to imagine alternatives to those that they experienced. (Goldspink & Foster, 2014, p. 151)

For the reasons outlined by Goldspink and Foster, policy making tends to be a relatively conservative activity. It is comparatively easy for curriculum traditionalists to appeal to shared experiences and memories with the aim of maintaining the status quo. It is much harder to explain unfamiliar new ideas that can't be succinctly explained and cannot appeal to widely shared experience. The TLRI project leaders were mindful of this risk because they were promoting large changes and unfamiliar curriculum directions. Wisely, they enlisted the support of powerful groups that could lobby in their support.

When expert professionals in a field say "this is really important", curriculum officials are likely to listen. Chris and Maxine were clear that the concept of bootstrapping was introduced to the senior secondary curriculum because the New Zealand Statistics

9 http://seniorsecondary.tki.org.nz/Mathematics-and-statistics

Association saw this as important and lobbied for its inclusion. I interviewed several members of this association, including Mike Camden, a long-time leader of the NZSA Education Committee. The interviewed NZSA members were also clear that they had played a successful role in supporting the development of the curriculum:

> Chris and Maxine are not alone. They have a lot of advocates around them.
>
> The committee [of the NZ Statistics Association] fostered good relationships with key Ministry of Education officials. We were very clear that we were speaking for the statistics community in New Zealand.
>
> The committee had a big influence on the shaping of the three strands of statistics in *NZC*.
>
> The big achievement was integrating statistical thinking into the curriculum. This has not been done elsewhere.
>
> The thinking is hard to ignore when all the pieces are integrated: curriculum, teaching and assessment tools, research, lobbying.

This section has outlined a range of policy-related actions that fit well with the concept of scholarship in practice. I suspect they would be seen as a much less comfortable fit with traditional ideas about scholarship. I was unaware of the strategic nature and extent of these collaborative activities before I conducted the interviews with key players inside the projects or in supporting roles close to the action. Looking back on everything I have documented about the work of this team it now seems to me that the activities outlined in this section have made the critically important difference to the impact of the research itself.

Creating overall conditions for successful knowledge-building in practice

It should be evident from the preceding sections that ongoing reflective critique was woven into the very fabric of the two TLRI projects featured in this report.

- Teachers and advisers were expected—and supported—to provide a robust critique of the practicality of the teaching and learning models being explored, and to identify any fishhooks or omissions in the policy instruments developed (both curriculum and assessment) and in the associated support materials.
- Professional statisticians were invited to provide critiques concerning the relevance and importance of the statistical thinking concepts and processes at the heart of the projects.
- The two project leaders challenged each other by bringing their different perspectives to bear on the decision making and actions taken as the project unfolded.

Although their contributions are listed separately here, all these groups came together in tightly networked "overlapping spheres of influence". Maxine sat at the very heart of these overlapping networks, holding them together, keeping ideas circulating and recruiting new members:

> You are building up networks all your life. (Maxine)
>
> Gravitate to people who are willing to engage with new ideas and push boundaries (Maxine)

The networks grew as teachers and advisers brought new recruits to the work, and also as the research leaders brought new researchers into the project team itself.

A quick study of Box 1 shows how many other researchers were also drawn into the work as it unfolded. Dr Stephanie Budgett, a member of the second project team, has now stepped up to be a co-leader of the third TLRI project,[10] and other team members are also showing research leadership in related areas. This is just one positive consequence of the strongly networked partnership model employed.

Reflective critique:
Clear connections are established between 'reasoning in action' (i.e. judgements made) and the consequences of the work. There is breadth and depth in the evidence presented to support this critique, and significant implications are drawn for future practice.

The role of the international community of statisticians and statistics educators

Statisticians and statistics educators from other nations are important groups whose contributions have not been specifically discussed in the report so far. Presenting at international conferences is a hallmark of traditional scholarship. Being an invited keynote speaker, or having a paper presentation accepted for a conference where places are strongly contested, is rightly regarded as evidence of peer esteem. But should we be looking for 'something more' than simply disseminating the research if our focus is on scholarship *in practice*?

The final criterion, as paraphrased above, directs attention to opportunities for robust critique that come from sharing thinking and new knowledge claims with a community of international peers. International symposia and conferences provide opportunities to

10 https://www.tlri.org.nz/tlri-research/research-progress/post-school-sector/visualising-chance-learning-probability-through

gain input from a wide range of well-informed and interested peers, and to compare and contrast New Zealand contexts and challenges with those that pertain elsewhere.

With this opportunity in mind, Box 3 summarises key interactions between the TLRI researchers and this wider international community as the two projects unfolded. Notice that some of these international events take place at regular intervals—typically every 2 years. Involvement in networking in this type of forum can be particularly valuable when strong relationships are already established and critical debates can develop and be revisited over time.

Other evidence of positive consequences

The final criterion for scholarship in practice challenges researchers to maintain ongoing scholarly attention to the *consequences* of their work. This is not easy when projects are funded in cycles and new projects are needed to renew these cycles of work. Again, this TLRI team has been strategic in making plans to keep looking for evidence of the impact of their work.

Box 3: International networking opportunities

2009: Chris gives the opening keynote presentation, and Maxine and Pip give a presentation, at the 6th Statistical Reasoning, Thinking and Literacy Forum in Brisbane, Australia.

Chris gives a keynote presentation at the US Conference on Teaching Statistics, Columbus, Ohio.

2010: Chris reads a paper at the Royal Statistical Society in London on 20 October 2010, World Statistics Day, during the launch of the UK 10-year statistical literacy campaign.

Maxine, Chris and another TLRI member give keynote presentations at the Fourth International CensusAtSchool Workshop, Royal Statistical Society, London, while Maxine and a another TLRI member also give two workshops.

Maxine, Chris and another TLRI member give keynote presentations at the GetStats week at Plymouth University, UK, while Maxine and a key TLRI member also give two workshops.

Maxine, Pip, and Chris each give a presentation at the 8th International Conference on Teaching Statistics in Ljubljana, Slovenia.

Chris gives a keynote presentation at the Australian Statistical Conference in Perth.

2011: Chris gives a keynote presentation at the Canadian Mathematical Society Conference, Toronto, Canada.

Maxine and another member of the TLRI team give a presentation at the 8th Delta Conference on the Teaching and Learning of Undergraduate Mathematics and Statistics, Rotorua.

Maxine gives a presentation at the 7th Statistical Reasoning, Thinking and Literacy Forum in Texel, The Netherlands.

Maxine gives a keynote presentation and workshop at the 51st Hungarian Annual Teachers Conference, Komarom, Hungary, and a presentation at Ludwigsburg University of Education, Germany.

At the time of writing this report, Maxine was mentoring a teacher partner from one of the TLRI projects. The teacher was undertaking a study of patterns of achievement in the NCEA achievement standard AS 91580 (Investigate Time Series Data) for an MSc thesis. If the two projects have indeed had a positive impact on students' statistical reasoning capabilities, and on teachers' pedagogical knowledge and skills to foster these capabilities, then strong achievement gains should become evident in related areas such as this. This work is not yet complete, but anecdotal evidence is indeed pointing to this type of positive impact. However, the more important point is that the team have not stopped looking just because the projects have been completed and they have moved on to new work.

Similarly, the wider team are continuing to monitor the consequences of the policy shifts they have helped bring about. When my report was almost complete I was shown a copy of the outcomes of a 2013 discussion forum sponsored by the NZ Statistical Association Education Committee. The forum had a specific focus on the nature of resources needed to support ongoing successful implementation of the new achievement standards. The memo included specific suggestions of strategies for meeting the needs the teachers had identified. The support efforts are clearly ongoing!

2012: Chris gives keynote presentations for the Int. Stat. Inst./KSS/KOSTAT Special Conference in Seoul, Korea; The Priestman Memorial Lectures 1 and 2, University of New Brunswick, Canada; and The Belz Memorial Lecture, the annual public Lecture of the Statistical Society of Australia's Victorian Branch.

Chris gives workshops in London to the Economic and Social Research Council of the UK, the Royal Statistical Society, the British Academy Conference; and also for University of Melbourne lecturers, the CensusAtSchool International Workshop in San Diego, USA, the University of Brunswick, Canada, and two high schools in Fredericton, New Brunswick.

Chris gives a CAUSE (Consortium for the Advancement of Undergraduate Statistics Education, USA) webinar presentation.

Maxine and another member of the TLRI team give a presentation to the International Association for Statistical Education Roundtable Conference in Cebu, the Philippines.

Maxine gives a Regular Lecture paper (keynote address) and another TLRI member gives a presentation at the 12th International Congress on Mathematics Education in Seoul, Korea.

2013: Maxine, Pip and another member of the TLRI team give a presentation at the 8th Statistical Reasoning, Thinking and Literacy Forum in Two Harbors, Minnesota, USA.

Maxine and another TLRI member each give a presentation at the 59th International Statistical Institute World Statistics Congress in Hong Kong. Chris gives a keynote presentation at the US Conference on Teaching Statistics in Raleigh, North Carolina.

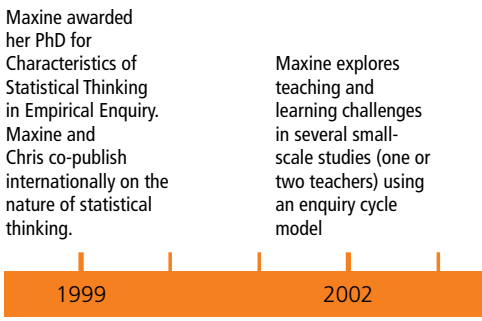
2014: Chris, Maxine, Pip, and two TLRI members each give a presentation at the 9th International Conference on Teaching Statistics at Flagstaff, Arizona, USA.

Concluding comments

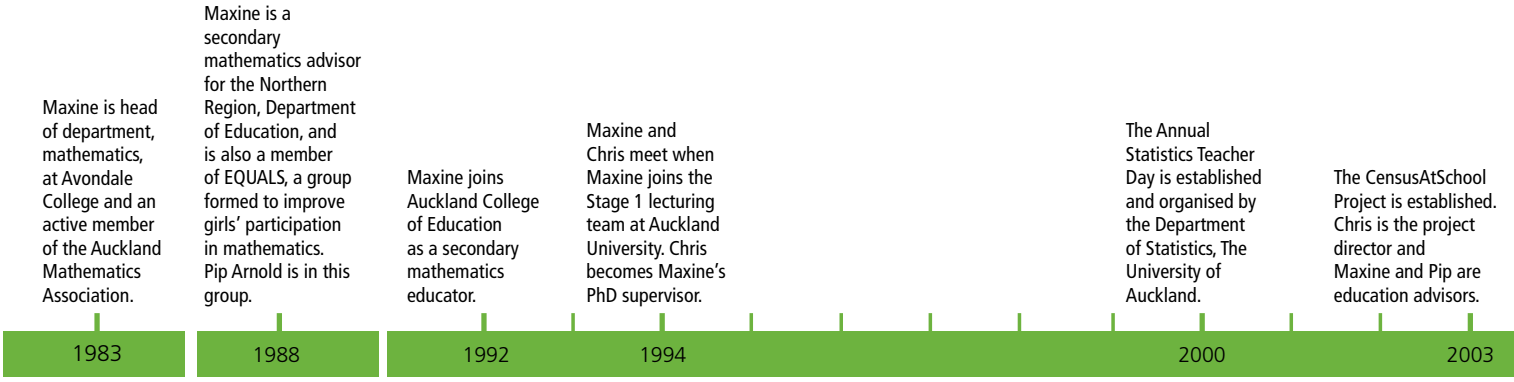
The projects featured in this first TLRI Project Plus report were chosen because they have been so successful in generating powerful new knowledge about New Zealand students' statistical reasoning, and for having a positive impact on the teaching and learning of that reasoning in our schools. My aim in developing this synthesis was to generate insights that might support further strong design work in future TLRI projects while celebrating the success of selected work already completed.

The concept of scholarship in practice has provided a strong theoretical framework within which to document and organise the complex nature of the activities that took place within the two featured projects. More than this, the six criteria for scholarship in practice have provided an explanatory framework that can account for the power and impact of these two projects. The model has been useful for interrogating judgement calls, such as when it is appropriate to develop resources as part of a project, how and why project participants might get involved in policy-related activities, and how best to leverage the partnership requirement so that everyone involved in the project learns something.

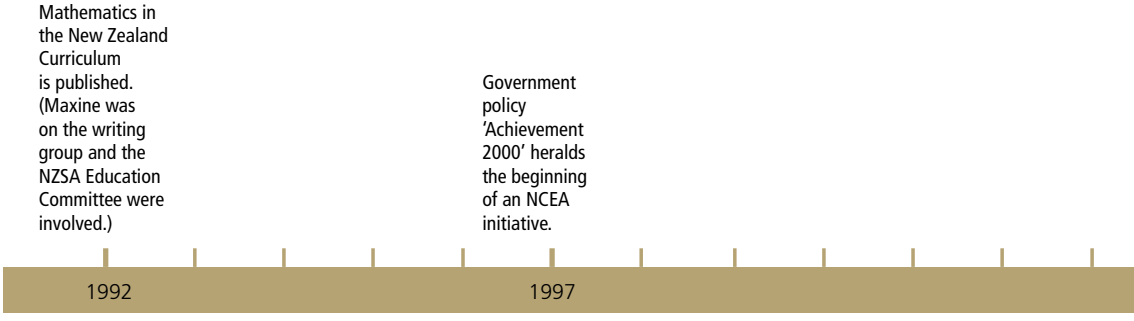
Timelines



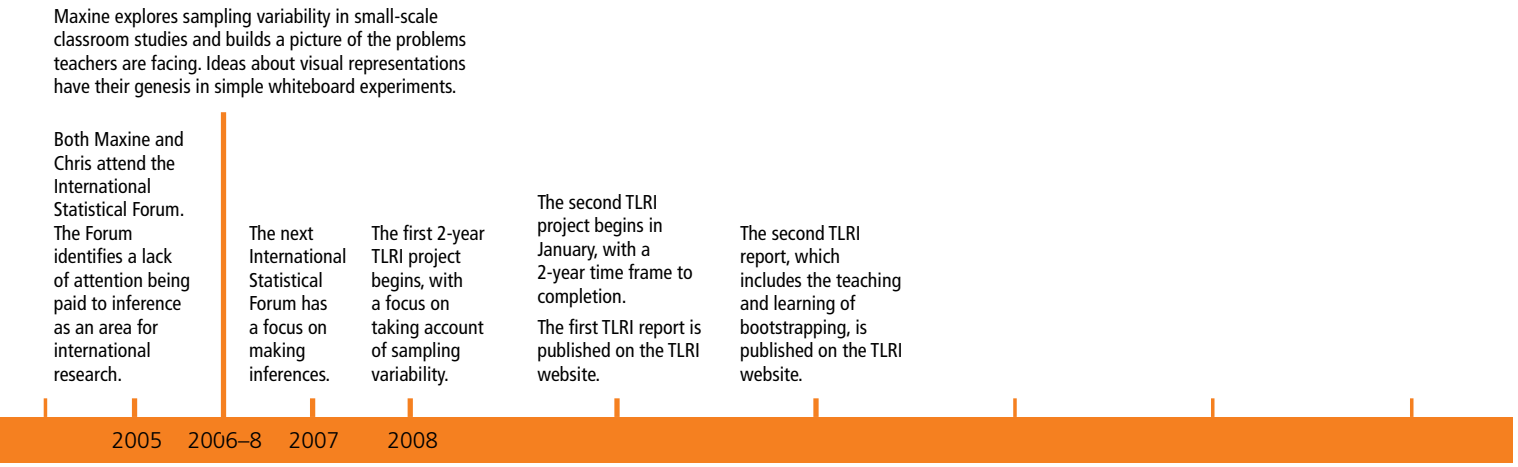
Research timeline

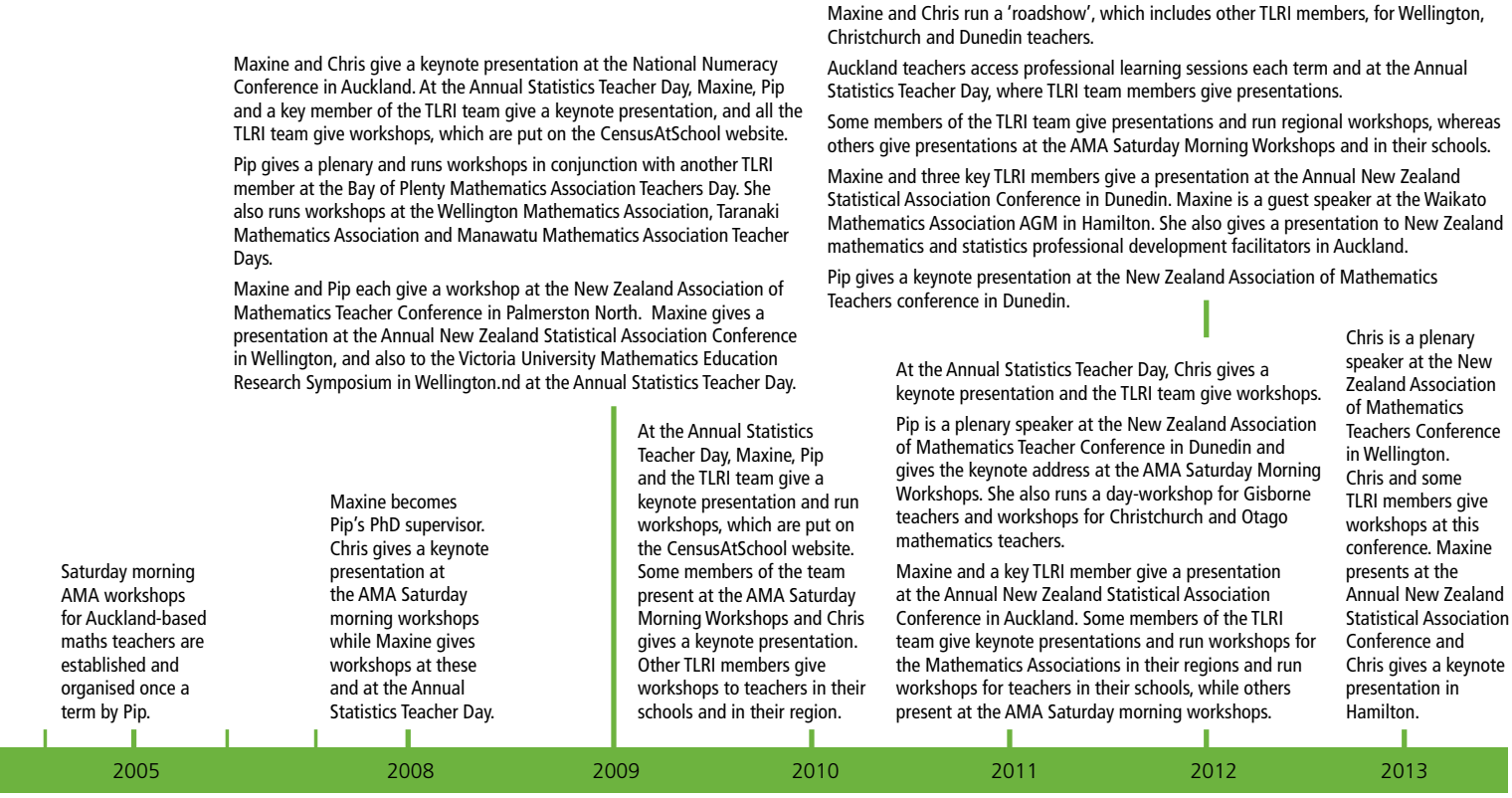


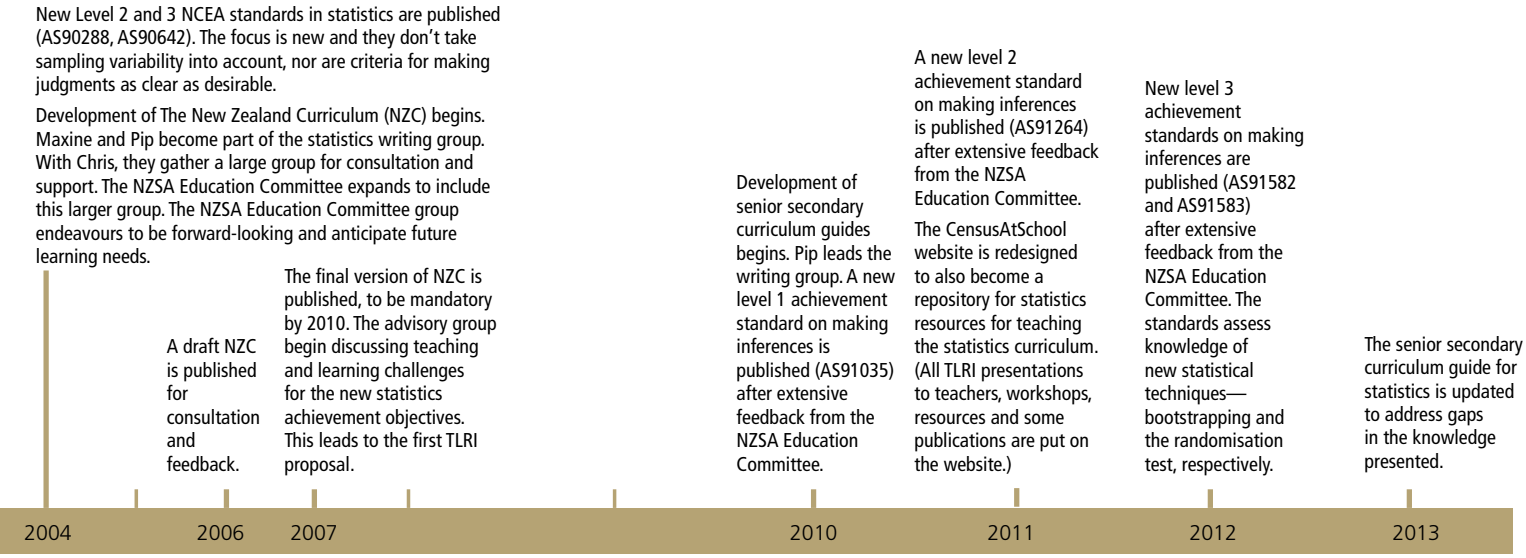
National networking timeline



Policy timeline


Maxine explores sampling variability in small-scale classroom studies and builds a picture of the problems teachers are facing. Ideas about visual representations have their genesis in simple whiteboard experiments.
Both Maxine and Chris attend the International Statistical Forum. The Forum identifies a lack of attention being paid to inference as an area for international research.
The next International Statistical Forum has a focus on making inferences.
The first 2-year TLRI project begins, with a focus on taking account of sampling variability.
The second TLRI project begins in January, with a 2-year time frame to completion.
The first TLRI report is published on the TLRI website.
The second TLRI report, which includes the teaching and learning of bootstrapping, is published on the TLRI website.
2005
2006–8
2007
2008


Maxine and Chris give a keynote presentation at the National Numeracy Conference in Auckland. At the Annual Statistics Teacher Day, Maxine, Pip and a key member of the TLRI team give a keynote presentation, and all the TLRI team give workshops, which are put on the CensusAtSchool website.
Pip gives a plenary and runs workshops in conjunction with another TLRI member at the Bay of Plenty Mathematics Association Teachers Day. She also runs workshops at the Wellington Mathematics Association, Taranaki Mathematics Association and Manawatu Mathematics Association Teacher Days.
Maxine and Pip each give a workshop at the New Zealand Association of Mathematics Teacher Conference in Palmerston North. Maxine gives a presentation at the Annual New Zealand Statistical Association Conference in Wellington, and also to the Victoria University Mathematics Education Research Symposium in Wellington.nd at the Annual Statistics Teacher Day.
Maxine and Chris run a 'roadshow', which includes other TLRI members, for Wellington, Christchurch and Dunedin teachers.
Auckland teachers access professional learning sessions each term and at the Annual Statistics Teacher Day, where TLRI team members give presentations.
Some members of the TLRI team give presentations and run regional workshops, whereas others give presentations at the AMA Saturday Morning Workshops and in their schools.
Maxine and three key TLRI members give a presentation at the Annual New Zealand Statistical Association Conference in Dunedin. Maxine is a guest speaker at the Waikato Mathematics Association AGM in Hamilton. She also gives a presentation to New Zealand mathematics and statistics professional development facilitators in Auckland.
Pip gives a keynote presentation at the New Zealand Association of Mathematics Teachers conference in Dunedin.
Saturday morning AMA workshops for Auckland-based maths teachers are established and organised once a term by Pip.
Maxine becomes Pip's PhD supervisor. Chris gives a keynote presentation at the AMA Saturday morning workshops while Maxine gives workshops at these and at the Annual Statistics Teacher Day.
At the Annual Statistics Teacher Day, Maxine, Pip and the TLRI team give a keynote presentation and run workshops, which are put on the CensusAtSchool website. Some members of the team present at the AMA Saturday Morning Workshops and Chris gives a keynote presentation. Other TLRI members give workshops to teachers in their schools and in their region.
At the Annual Statistics Teacher Day, Chris gives a keynote presentation and the TLRI team give workshops.
Pip is a plenary speaker at the New Zealand Association of Mathematics Teacher Conference in Dunedin and gives the keynote address at the AMA Saturday Morning Workshops. She also runs a day-workshop for Gisborne teachers and workshops for Christchurch and Otago mathematics teachers.
Maxine and a key TLRI member give a presentation at the Annual New Zealand Statistical Association Conference in Auckland. Some members of the TLRI team give keynote presentations and run workshops for the Mathematics Associations in their regions and run workshops for teachers in their schools, while others present at the AMA Saturday morning workshops.
Chris is a plenary speaker at the New Zealand Association of Mathematics Teachers Conference in Wellington. Chris and some TLRI members give workshops at this conference. Maxine presents at the Annual New Zealand Statistical Association Conference and Chris gives a keynote presentation in Hamilton.
2005
2008
2009
2010
2011
2012
2013


New Level 2 and 3 NCEA standards in statistics are published (AS90288, AS90642). The focus is new and they don't take sampling variability into account, nor are criteria for making judgments as clear as desirable.
Development of The New Zealand Curriculum (NZC) begins. Maxine and Pip become part of the statistics writing group. With Chris, they gather a large group for consultation and support. The NZSA Education Committee expands to include this larger group. The NZSA Education Committee group endeavours to be forward-looking and anticipate future learning needs.
A draft NZC is published for consultation and feedback.
The final version of NZC is published, to be mandatory by 2010. The advisory group begin discussing teaching and learning challenges for the new statistics achievement objectives. This leads to the first TLRI proposal.
Development of senior secondary curriculum guides begins. Pip leads the writing group. A new level 1 achievement standard on making inferences is published (AS91035) after extensive feedback from the NZSA Education Committee.
A new level 2 achievement standard on making inferences is published (AS91264) after extensive feedback from the NZSA Education Committee.
The CensusAtSchool website is redesigned to also become a repository for statistics resources for teaching the statistics curriculum. (All TLRI presentations to teachers, workshops, resources and some publications are put on the website.)
New level 3 achievement standards on making inferences are published (AS91582 and AS91583) after extensive feedback from the NZSA Education Committee. The standards assess knowledge of new statistical techniques—bootstrapping and the randomisation test, respectively.
The senior secondary curriculum guide for statistics is updated to address gaps in the knowledge presented.
2004
2006
2007
2010
2011
2012
2013

References

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

Goldspink, C., & Foster, M. (2014). Learning to learn at a whole system level: Development of the South Australian Teaching for Effective Learning Framework. In R. Deakin-Crick, C. Stringher & K. Ren (Eds.), *Learning to learn: International perspectives from theory and practice* (pp. 145-169). London, UK: Routledge.

Hipkins, R., Bolstad, R., Boyd, S., & McDowall, S. (2014). *Key competencies for the future* Wellington: NZCER Press.

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

Penuel, W., Coburn, C., & Gallagher, D. (2013). Negotiating problems of practice in research–practice design partnerships. *National Society for the Study of Education*, *112*(2), 237–255.

Singer-Gabella, M. (2012). Toward scholarship in practice. *Teachers College Record*, *114*(8), 1–30.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher learning and professional development: BES*. Wellington: Ministry of Education.

Wild, C., Pfannkuch, M., Regan, M., & Parsonage, R. (2013). *Next steps in accessible conceptions of statistical inference: Pulling ourselves up by the bootstraps*. Retrieved from https://www.stat.auckland.ac.nz/~wild/TEMP/bootstrap.pdf

Acknowledgments

I am grateful for the time that various members of the TLRI statistical reasoning team gave to this, the very first TLRI Project Plus initiative. In face to face conversations and via email we kept up a lively correspondence as I worked to build an accurate account of their complex and tightly networked activities. Maxine Pfannkuck, in particular, facilitated and supported my work every step of the way.

This project was made possible by the vision and strategic thinking of those who lead the management of the TLRI fund, in particular NZCER Director Robyn Baker. Working with Robyn, researcher Josie Roberts first picked up the challenge and developed the brief for the work. Jo MacDonald, who manages the fund overall, has been a steady support and critical eye as the project unfolded. NZCER's publisher, David Ellis, managed the publication process.