STATISTICS & DATA SCIENCE EDUCATOR

What's in a Pasifika name? Constructing a name dataset

Malia S. Puloka and Maxine Pfannkuch
Waipapa Taumata Rau
The University of Auckland, New Zealand
Email: m.puloka@auckland.ac.nz

Overview of lesson

The lesson introduces Year 9 students to constructing variables and designing survey questions for categorical data through the context of their names. In a teacher-supported lesson students are encouraged to consider characteristics of names to form variables in preparation for designing survey questions, followed by data collection, and interpreting the resultant datasets. The activity engages students in the first three phases of the PPDAC cycle.

Learning objectives

- Construct variables for categorical data
- Design survey questions
- Read datasets

Suggested age range

12-15-year-olds

Time required

Approximately 100 minutes

Keywords

categorical data, constructing variables, designing survey questions, interpreting datasets

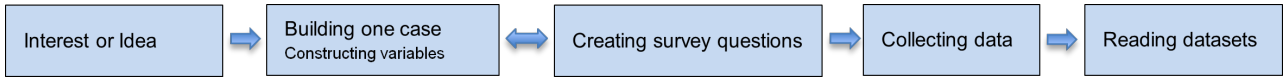
Introduction

I, the first author, designed and implemented this activity as part of my PhD research into posing and answering questions related to categorical data. A pre-test, ten-lesson teaching intervention, and post-test were conducted with 13-14-year-old mixed ability Pasifika and Māori students in a low-socio-economic school. The pre-test revealed the students were not familiar with reading and interpreting datasets (Puloka et al., 2021). Because the students had no experience of constructing variables, defining measures, designing survey questions, participating in surveys such as CensusAtSchool, or reading and interpreting data structured in a spreadsheet, there was a need to introduce them to the first three phases of the PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle for statistical investigations. Constructing variables and defining measures are human-based decisions and need urgent attention in teaching and research, particularly in today's data-drenched and AI world (Arnold, 2022). Furthermore, how variables are created, and measures are defined are critical "worry" questions when evaluating a study (Budgett & Pfannkuch, 2010) and can lead to biases in predictive modelling. When considering the context for constructing variables, I chose to use student names, as great importance is attached to naming a child in Pasifika communities and thus, I could integrate and weave the students' learning into their culture and interests.

Photo: Malia S. Puloka

Lesson sequence outline

The aim of this three-lesson sequence was to create a dataset using the following structure:



The sequence traversed three lessons. Setting the scene, building one case, and creating survey questions occurred in the latter half of the first lesson. In the second lesson building one case and creating survey questions was revisited, as the act of creating survey questions stimulated the students to think of more variables, and then the data was collected. At the beginning of the third lesson the students received the data collected.

1. Interest or idea (Lesson 1)

Since all the students were of Pasifika or Māori ethnicity, 'Name' was chosen as a topic of interest because of its relationship to their culture and, at the same time, Names have special and interesting meanings that young people do not often think about. Traditionally in the Tongan culture when a child is born, especially if they are firstborn and even more so if it is a son, the family of the baby's father has the honour and responsibility of naming the child. This is usually the paternal grandfather's responsibility. Sometimes, the honour is given to the eldest sister of the paternal grandfather who has a special rank in the family hierarchy. The name given to the newborn is usually a namesake in order for that child to carry a name that is unique or special to the family such as a chief's name. The name given could also be from someone with a special status in the family such as one from the line of fathers. It is also common to name someone based on a significant event or place. The family of the baby's mother is rarely given the honour of naming the child. However, the expectations and responsibility of naming a child are not rules that must be followed. In some cases, the parents have a say in what they want to call their child or to whom they want to give the naming honour.

To lead students to think about the importance of names and to stir interest, I discussed the names of popular international rugby union teams. Since most of the students in the class were boys and rugby union is the number one sport in New Zealand, names of teams seemed to be suitable. I began by asking the students whether they knew the meaning of the names of rugby union teams, such as "All Blacks", the New Zealand rugby union team, and "Springboks", the South Africa rugby union team. This started a conversation between me and the students. I then asked the students whether they knew the meaning of their names and who named them. Since there was no volunteer, I started with my name, Malia, and that I was named after the mother of Christ. The following exchange took place.

S: I thought it's Mary.

S: It is Mary.

T: It is Mary. Malia in Tongan.

S: Malia?

S: Mele is Tongan.

T: or Mele.

Mele is a very common Tongan name and the common translation for 'Mary'. Only the Catholics use 'Malia' for the mother of Christ. The other churches use 'Mele'. That explains why the students brought up 'Mele' for 'Mary'.

Before thinking about questions to investigate, it is important to understand and build up one's contextual knowledge of the situation (Wild & Pfannkuch, 1999). To build up students' contextual knowledge of names I chose to focus on my name to stimulate students to think about what socio-cultural history might lie behind it and at the same time reflect on their own names. Looking at one case, my name, also had the additional benefit of keeping students focussed.

2. Building one case and creating the survey questions (Lesson 1)

The conversation then moved to a discussion about some of the characteristics of the name Malia, and we started to co-construct variables, building up a diagram in the process (Figure 1). Because I was named Malia, a religious name, we agreed that

'religious' was a characteristic of the name (see Figure 1). I asked, *What else can you say about my name*? A student responded, *Has a 'm'. There's 'a', there's a 'l', there's a 'i', there's a 'a'*, so the characteristic 'starts with M' was added. In response to a student noticing that my name ended with an 'a', I asked, *What do you notice about Pasifika and Māori names, the traditional ones, not the pālangi ones?* Student responses included, *they sound the same, it always ends in a vowel*, and after further discussion we added the characteristic, '3 vowels', to the diagram. To stimulate the students to think further about the difference between Pasifika names and pālangi names I asked if they heard the names of two people were Jamie and Semisi, what would they think. They responded that Jamie was pālangi and Semisi was Tongan, and therefore I suggested that a characteristic could be 'ethnic origin' of name. The 'female' characteristic was added after I asked if anyone had a unisex name, a name that could be a boy's or girl's name. The final two characteristics, 'place of birth' and 'namesake', were agreed upon through students asking questions about me and my name.

T: What else do you want to know about me?

S: What village are you from?

T: Vava'u, Neiafu.

S: Miss, how many people were you named after?

T: Named after? Or whether my name is a namesake? Yes, I think my name is a namesake cos I'm named after the mother of Christ.

To end the lesson, I pointed to several characteristics of my name and asked, *What (survey) question would you ask for this one?* I wrote their responses on the whiteboard (see Figure 1).

The idea of constructing variables and noticing, wondering, and asking questions, in general, was new to the students as well as thinking about the rich array of variables and subsequent survey questions that could be derived from one's name. Therefore, I decided to revisit my name in the next lesson to stimulate students to think more deeply about its characteristics. It was also noticeable that discussion about names engaged the students as they became more attentive as the lesson progressed.

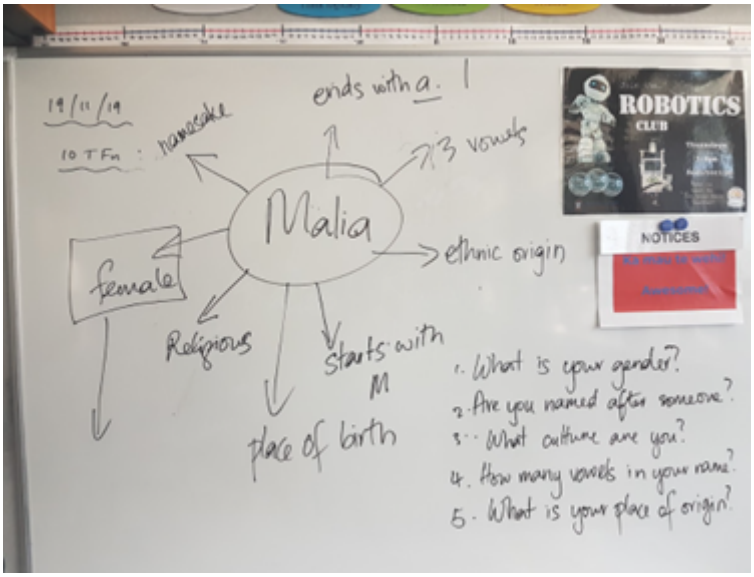


Figure 1: Lesson 1: Diagram of characteristics of the name Malia with survey questions alongside

3. Building one case and creating the survey questions revisited (Lesson 2)

To start the lesson, I said we were going to think about my name again and what else they could add or want to know about a name. The following interchange occurred when I asked students whether their name was long or short.

S: Long

S: It's a difficult name.

S: Short.

T: A difficult name? You find that others are finding it difficult to pronounce your name? Is it because it's long or is it because it's a Tongan name? How would you decide if a name is long or short name?

S: The number of letters.

T: The number of letters? So, do you think my one is long or short?

S: Short.

T: So, what would be something that is long?

S: His name.

T: Vahafolau's name? So, we can decide. My name has five letters and according to you guys this is short. So how many letters before you start saying that is a long name?

S: After 10. Five. Eight. (A few other students said eight.)

T: Eight? Who has eight or more letters in their name? (Vaha put up his hand.) Vahafolau. So, nine letters. So, this is long. So, do you all agree that something like eight or more letters is long?

S: Yeah.

Through this discussion we co-constructed the variable, length of name, and agreed to define a long name as one with eight or more letters and a short one as one with 1-7 letters (see Figure 2). Therefore, students were learning about the need to define a measure to be used and that it was based on their or a human decision. After this discussion I asked what else did they want to know about my name, to which one student responded, *Who named you?* I said my grandfather named me and as I was adding this characteristic to my name on the whiteboard a student remarked that the tradition of naming was interesting.

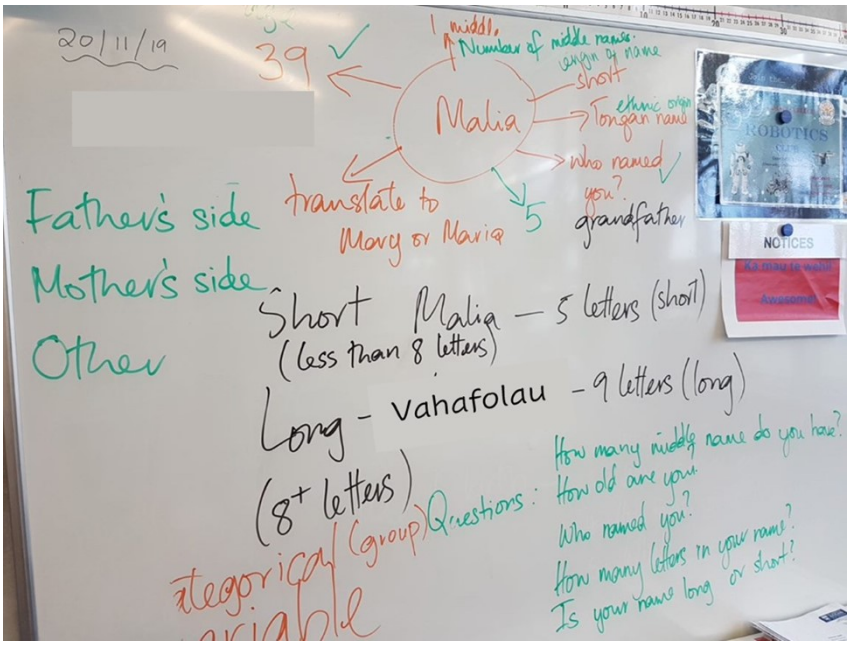


Figure 2: Lesson 2: Diagram with more characteristics for the name Malia with further survey questions alongside

The next question was my age, which was added as a characteristic.

> Age can provide an interesting discussion, for example, on how one might be able to tell the age of someone based on their name as names tend to increase and decrease in popularity throughout the decades (e.g., see: baby name visualizer and baby name popularity over time and other websites)

We then turned to creating some more survey questions from the characteristics written on the whiteboard. The students responded with, for example, *What is your age? How many vowels in your name? How many letters?* (see Figure 2). However, when they got to the characteristic short, they had some difficulty with one student, for example, suggesting *What is your height?* After I queried her response, she suggested *How long is your name?* We reached an impasse in the discussion so I suggested *Is your name long or short?*, which students agreed was an appropriate question. During further discussion a student came up with idea of asking about middle names because some people have many middle names because of their culture. I said I had one middle name and asked students to put their hands up if they had more than one middle name with their responses including *Three, Two,* and *I got a long one*. Because the students insisted on knowing my middle name, I said it was Sālote, which a student immediately associated with the Queen Sālote school in Tonga.

> Using my name as a context to create survey questions not only helped students to realise how one could construct variables through deconstructing the features of a name (e.g., number of vowels) and through drawing on the socio-cultural history behind one's name but also allowed them to find out more about me, a useful pedagogical practice for setting up a relationship with a class (Averill, 2012). Revisiting this activity seemed to suit their style of learning as the class had a short attention span. They also engaged well during our interactive dialogue. From the student responses it was evident that creating variables and survey questions needed many more lessons to improve and enhance their understanding of this important area of statistics.

4. Collecting data (Lesson 2)

During the lesson I typed up the survey questions (Figure 3) and provided a pre-prepared survey card for each student on which to put their responses to the survey questions.

Question	Variables	Survey questions
1	Name	What is your preferred name?
2	Gender	What is your gender? Male/Female
3	Ethnicity	What is your ethnicity?
4	Age	How old are you?
5	Starting letter	Does your name start with a consonant or a vowel? Consonant/Vowel
6	Number of letters	How many letters in your name?
7	Number of vowels	How many vowels in your name?
8	Length of name	Is your name long or short? Short (1 - 7 letters) / Long (8 or more letters)
9	Middle names	How many middle names do you have?
10	Namesake	Is your name a namesake? Yes / No / Don't know
11	Origin	Where did your name originate from? Father / Mother / Other
12	Name from	Where does your name come from? (e.g., a Tongan name, an English name)
13	Unisex	Is your name a unisex name? Yes / No / Don't know
14	Born	What is your place of birth?
15	Side of family	Whose side of the family named you? Father / Mother / Other

Figure 3: Name data variables and survey questions

The survey questions were projected one at a time onto a screen. As they filled out their survey card there were questions from the class as they sought to clarify either what the question meant or how they should answer a question. For example, Question 1 was, 'What is your preferred name?' When we got to Question 6, 'How many letters in your name? one of the students wanted to make sure that Question 6 was about the name they put down in Question 1. For Question 10, 'Is your name a namesake?' one student asked, *What if you're like a son of someone with your name?* This student used the name 'Junior' in Question 1. Because I understood that the student had the same name as his father, but he is called 'Junior', I told him that his name is a namesake. Where students did not know the answer to a question, such as where their name came from, I said to put down 'I don't know'.

Figure 4 shows a filled Name survey card with the name of the student in cell 1 hidden to keep their identity anonymous for ethical reasons. Where a student put more than one ethnicity, only the one they most identified with was chosen to be recorded in the final dataset.

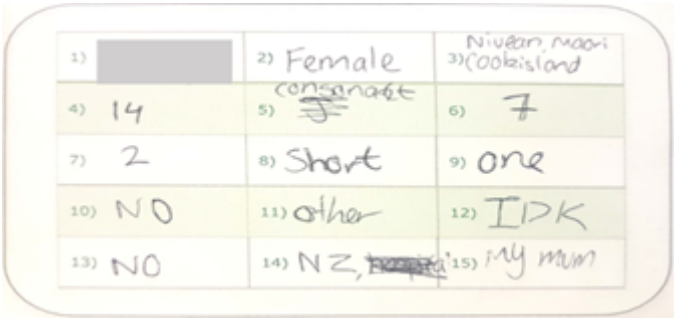


Figure 4: Example of Name survey card completed by a student

Survey cards were used to ensure that students would be familiar with this format because they would be using them extensively as data cards for learning about displaying and interpreting categorical data. Projecting the survey questions one at time worked well as it allowed the students to make sure they understood the question and could answer with confidence. On reflection the wording of the survey questions needed to be improved and gave rise to anomalies and ambiguities. For example, some survey questions were being answered about the preferred name and some questions about their actual names leading to a survey questionnaire that was internally inconsistent.

5. Reading datasets (Lesson 3)

Before the third lesson, I created the dataset (Figure 5), which I put on an A4 sheet to hand out to the students. On the back of the dataset sheet, I put the survey questions as a reminder and because some students were absent when the data was collected.

I also created sets of data cards for students to use in pairs. A data card is shown in Figure 6 with the name of the student replaced with XXXX for confidentiality reasons.

I handed out the dataset sheet to the students and they immediately sought to find themselves in the dataset and then proceeded to look at information about other students sitting near them. Some information was surprising to them such as one Tongan student discovering that his friend was born in Tonga not New Zealand as he was. Another student was surprised his friend was still 13 and asked him when his birthday was. They also checked for similarities between themselves such as wondering whether their friends had one middle name like them. Two friends shook hands because they were very delighted to find out that they both had the longest names in the class.

name	gender	ethnicity	age	s_letter	numletters	vowels	length	middlenam	namesake	origin	namefrom	unisex	born	who_side
	male	Maori	13	cons	4	2	short	2	yes	father	English	don't know	NZ	father
	male	Samoan	14	cons	5	1	short	1	yes	mother	Samoan	yes	NZ	mother
	male	Fijian	14	cons	5	2	short	2	yes	mother	Fijian	no	NZ	mother
	female	Maori	13	cons	5	2	short	1	no	other	don't know	no	NZ	mother
	male	CI Maori	13	cons	6	3	short	2	yes	father	Cook Is	no	Cook Is	father
	male	Tongan	14	cons	6	3	short	0	yes	father	Tongan	no	NZ	father
	female	CI Maori	14	cons	7	2	short	1	no	other	don't know	no	NZ	mother
	male	CI Maori	14	cons	7	3	short	1	don't know	mother	don't know	don't know	NZ	father
	male	Samoan	13	cons	8	4	long	0	don't know	father	Samoan	no	NZ	father
	male	Tongan	14	cons	9	3	long	2	no	father	Tongan	no	Tonga	father
	male	Tongan	14	cons	9	5	long	1	yes	father	Tongan	no	NZ	father

Figure 5: The Name dataset with students' names removed for confidentiality reasons

XXXX	female	CI Maori
14	cons	7
2	short	1
no	other	don't know
no	NZ	mother

Figure 6: An example of one of the data cards from the set of data cards given to students

After the students had looked at the dataset, I checked to make sure they understood the meaning of some of the variables. Confident that the students seemed to be able to read and interpret the dataset, I then handed out sets of data cards to each student pair and asked them to tell me about a particular name on a data card. Below is an example.

T: J and C, can you tell me about this card over here.

J: He's XXXX.

C: He's a male, he's Fijian, he's 14 years old, umm

J: His name starts with a consonant; he's got five letters in his name.

Together: He's got two vowels.

C: His name is short. What's this? (Points to the card.) Is this middle name?

J: Yes.

C: And is this namesake? Yes.

J: He's come from his mother's side. It's Fijian, his name–it's not unisex; he was born in NZ; he's come from his mother.

Once the students knew what they were expected to do, they each chose a card and explained it to each other.

Yeah. Her name is XXXX, she's a female, she's Cook Is Māori, she's 14, her name starts with a consonant, she has seven letters in her name, she has two last names, her name is short. (He paused, realising that he did not know what the 2 stood for. He checked the survey questions on the back of the dataset sheet.) Ooh, she got two vowels in her name. Her name is not a namesake.

Often, they asked each other to do their name. Note also in this next excerpt how their articulation of the who_side variable needed to be corrected, which I did not do in the example concerning J and C.

S1: His name is XXXX, he's a male, he comes from Tonga, he's 14, his name starts with a consonant; he has nine letters in his name; he has three vowels, his name is long, got two middle names, it's not a namesake, it comes from his mother?

S2: Father

S1: Ooh father. It's a Tongan name, it's not a unisex, he was born in Tonga, he comes from his father.

Because his interpretation of the who_side variable was ambiguous and was the same as the origin variable I asked him to look at the survey question and clarified that it was about whose side of the family named the student.

On reflection, three of the survey questions (11, 12 and 15, see Figure 3) needed to be reworded and improved as the meaning of the variables was not clear to the students. Also, the students and I needed to articulate and communicate more clearly and precisely the data and meaning of all the variables and particularly for the three variables in questions 11, 12 and 15 (origin, namefrom, who_side).

For the above excerpt, the phrases, he comes from Tonga, he has three vowels, it comes from his father, and he comes from his father, were referring to the variables, ethnicity, vowels, origin and who_side respectively. Precise communications for the meaning of the four variables and the data from the case, respectively, could be: his ethnicity is Tongan; his name has three vowels; the name comes from the father's side of the family; and the father's side of the family named him.

The students were engaged as they seemed interested in finding out more about their classmates. They helped each other to read the data cards and referred to the survey questions when they were unsure of the meaning of the information. Learning to read the dataset and data cards and to connect the information to the survey questions is an essential prerequisite before starting a statistical investigation (Arnold, 2022).

Because my PhD research was focused on posing and answering questions related to categorical data, the dataset and data cards were then used in many of the activities during the teaching intervention.

Adaptations

The following adaptations are suggested to enrich this activity depending on the learning stage of the students and the learning sequence being designed.

- Ethics are foundational to data collection and are now prominent in the New Zealand refreshed curriculum across all curriculum levels (Ministry of Education, 2023). When collecting data from a class, consider and discuss ethics with the class before deciding what data to collect. For the name survey, teachers would need to be mindful of the background of students including being sensitive to participants' cultural and religious perspectives and whether sharing any of their name information with the class might not be appropriate. For a full discussion on ethics see Arnold (2022, Section 3.3, p. 95).

- Arnold (2022, p. 79) devised 12 questions to interrogate a dataset. To reinforce the connection between the dataset/data cards and the survey questions, Arnold's questions 7 to 12 would be useful for students to answer such as *State the variable*, *What was the survey question asked?* and *What are possible outcomes for the variable?*

- After data collection get students to reflect on, evaluate and critique the survey questions. For example: *Could students interpret some questions differently? If they gave the survey questionnaire to another class, would any of the survey questions be confusing (because they were not involved in the design of the survey questions)? Are the survey questions relevant to the area of interest? Why or why not? Will the survey questions enable them to learn more about names when they investigate and explore the data? What other survey questions could be asked? What wording would improve some survey questions? Why?*

- Students can search the internet to find out when names were popular, the meaning of

names and other information to produce more survey questions to enrich their investigations about names.

Teacher notes

The activity, based around names, drew on Pasifika knowledge, culture, identity, and context and is an example of how the Pacific values of Alofa (Love, Dignity, Respect) and Fonua (Belonging, Connectedness) (Ministry of Education, 2020a; 2020b) can be incorporated into teaching and learning statistics.

In terms of the PPDAC cycle the process can be thought of as PPD–PAC (Figure 7), where this activity covers PPD and once the data are collected the next lessons can cover PAC, the statistical investigative questions followed by analyses and conclusions. Because my focus was on categorical data, the investigative questions in the follow up activities covered relationship questions for paired categorical data: simple proportion, joint proportion, conditional proportion, and comparison situations about a group.

From problem to plan, the process used in this activity is captured in Figure 8, where a broad interest in names gradually becomes encapsulated into precise survey questions. Contextual knowledge drives the earlier stages with statistical knowledge contributing more to the latter stages.

To resolve the problem of students not articulating the meaning of the variables and data precisely, a good pedagogical practice would be to foster communication. The work of Arnold and Pfannkuch (2022) described how a teacher improved students' descriptions of data visualisations and Arnold (2022) also described how to improve students' ability to pose good investigative questions, whereby students practised these skills at the start of each lesson. The teacher is central to fostering communication through *knowing* how to articulate precisely, verbally and in writing, the meaning of variables, *modelling* use of the language required, and actively *listening* to student articulations and identifying, with student contributions, what is good and what needs to be improved. Figure 9, based on the work of Arnold and Pfannkuch (2022) proposes a potential avenue for teachers to explore how to improve student communication of the meaning of variables.

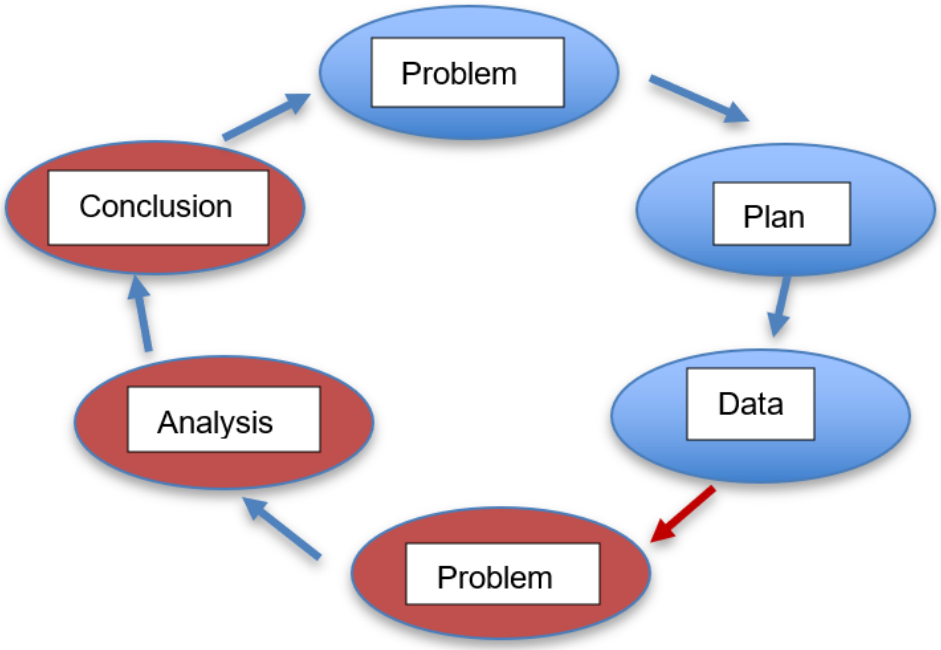
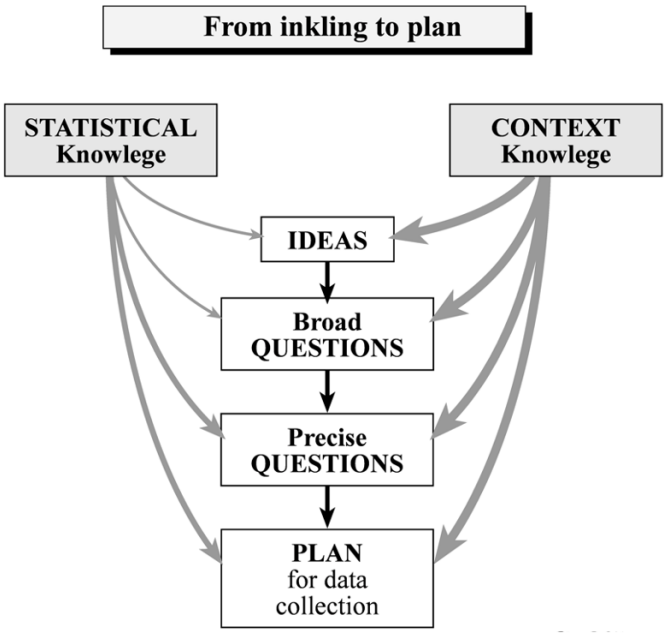


Figure 7: Statistical investigation cycle used for teaching intervention, with only PPD described in the activity



Commentary on inkling to plan

Note the width of the arrows indicates the input required from the contextual and the statistical knowledge domains.

IDEAS: Draw on socio-cultural history about names as contextual knowledge to garner ideas about my name. Notice differences and similarities among names (e.g., length). At the same time use statistical knowledge to inform me about what ideas may lead to data that would be possible to collect.

BROAD QUESTIONS: Turn ideas about my name into characteristics of my name, the beginnings of variable identification and definition.

PRECISE QUESTIONS: Turn characteristics of my name into survey questions that can be answered.

PLAN: Collect data from class by projecting survey questions onto whiteboard with students recording their responses onto a data card.

Figure 8: Pictorial description of PPD in the activity. Diagram source from Wild and Pfannkuch (1999, p. 228), which is reinterpreted for this activity

name	gender	ethnicity	age	is_letter	numletters	vowels	length	middlenam	namesake	origin	namefrom	unisex	born	who_side
	male	Maori	13	cons	4	2	short	2	yes	father	English	don't know	NZ	father
	male	Samoan	14	cons	5	1	short	1	yes	mother	Samoan	yes	NZ	mother
	male	Fijian	14	cons	5	2	short	2	yes	mother	Fijian	no	NZ	mother
	female	Maori	13	cons	5	2	short	1	no	other	don't know	no	NZ	mother
	male	CI Maori	13	cons	6	3	short	2	yes	father	Cook Is	no	Cook Is	father
	male	Tongan	14	cons	6	3	short	0	yes	father	Tongan	no	NZ	father
	female	CI Maori	14	cons	7	2	short	1	no	other	don't know	no	NZ	mother
	male	CI Maori	14	cons	7	3	short	1	don't know	mother	don't know	don't know	NZ	father
	male	Samoan	13	cons	8	4	long	0	don't know	father	Samoan	no	NZ	father
	male	Tongan	14	cons	9	3	long	2	no	father	Tongan	no	Tonga	father
	male	Tongan	14	cons	9	5	long	1	yes	father	Tongan	no	NZ	father

XXXX	female	CI Maori
14	cons	7
2	short	1
no	other	don't know
no	NZ	mother

Datasets	Teacher Role	Possible dialogue
	The teacher **knows** how to articulate the meaning of the variables and data. The teacher **models** verbal and written descriptions using student ideas. The teacher, together with the students, update student language to precise language. The teacher **listens** to student contributions. What is present and what is missing is identified by the teacher and students.	S: She comes from her mother (referring to mother, bottom right on data card). T: What does "she" refer to? S: Her name. T: Does the variable mean she is named after her mother? S: No T: Does the variable mean her mother named her? S: No T: Can anyone clarify the variable for the statement we now have, 'Her name comes from her mother'. S: The mother's side of her family named her?

Figure 9: Role of teacher in improving student communication. Adapted from Arnold and Pfannkuch (2022, p. 7)

For more information on designing good survey questions and how survey questions can be refined over time see Arnold (2022, pp. 315–318) and see the CensusAtSchool teacher's guide to learn about the information that lies behind some of the CensusAtSchool survey questions.

References

Arnold, P. (2022). *Statistical investigations | Te Tūhuratanga Tauanga*: *Understanding progressions in The New Zealand curriculum and Te Marautanga o Aotearoa.* NZCER Press.

Arnold, P. & Pfannkuch, M. (2022). Engaging novice statisticians in statistical communications. *Mathematics Education Research Journal*. https://doi.org/10.1007/s13394-022-00442-w

Averill, R. (2012). Caring teaching practices in multiethnic mathematics classrooms: Attending to health and well-being. *Mathematics Education Research Journal, 24*(2), 105–128. https://doi.org/10.1007/s13394-011-0028-x

Budgett, S., & Pfannkuch, M. (2010). Assessing students' statistical literacy. In P. Bidgood, N. Hunt & F. Jolliffe (Eds), *Assessment methods in statistical education: An international perspective* (pp. 103–121). John Wiley & Sons Ltd.

Ministry of Education (2020a). Pacific values framework: Designing for Pacific learners and contexts. Author.

Ministry of Education (2020b). Pacific values framework: Learning area guides. Author.

Ministry of Education (2023). *Te Mātaiaho |The refreshed New Zealand Curriculum: Mathematics and Statistics.* Author.

Puloka, M., Budgett, S., & Pfannkuch, M. (2021). What questions do novices pose about categorical data? In R. Helenius & E. Falck (Eds.), *Statistics Education in the Era of Data Science Proceedings of the Satellite conference of the International Association for Statistical Education (IASE), Aug-Sept 2021, Online conference*. https://doi.org/10.52041/iase.lciru

Wild, C., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry (with discussion) *International Statistical Review 67*(3), 223–265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Materials required

To make data cards automatically from datasets see: making your own datacards.

Copyright information