Fourth-Graders' Meta-questioning in Statistical Investigations

Lyn D. English
Queensland University of Technology
<l.english@qut.edu.au>

Jane M. Watson
University of Tasmania
<jane.watson@utas.edu.au>

Noleine Fitzallen
University of Tasmania
<noleine.fitzallen@utas.edu.au>

This paper addresses the initial components of an activity in which 4th-grade students engaged in meta-questioning as they created and refined survey questions with the aim of comparing life across two Australian cities. We propose the term, *meta-questioning*, as a core, underrepresented feature of statistical investigations in the primary school. We report on the nature of the students' initial posed questions and their subsequent refined questions, students' justifications for their question refinements, their anticipated data collection, and developments in their question posing skills. Results include a hierarchy of question types posed by the students and how their question types changed with subsequent refinements.

A statistical question is the starting point for any investigation. Yet, posing questions is underrepresented in many curriculum documents (Allmond & Makar, 2010; English & Lehrer, in press; Lavigne & Lajoie, 2007), perhaps because initial statistical questions are often informal and broad and require substantial refining (English, 2014; Whitin & Whitin, 2011). Young students often find it difficult to generate questions that can be investigated or to envision the data that can address their questions (Allmond & Makar, 2010). Lehrer and Schauble (2002) suggested that many of the challenges of posing questions for children can be ameliorated when they are given opportunities to build familiarity with the phenomena being investigated, including talking about and observing the target phenomena. Encouraging young students to evaluate questions posed critically is a foundational statistical process that is often overlooked in mathematics curricula. Yet an ability to deal intelligently with the myriad data in everyday events, including questioning data sources and questions investigated, is essential if citizens are to engage effectively in democratic discourse and public decision-making.

This paper addresses a core component of statistical investigations in the primary school, namely, students' developments in meta-questioning (adapted from Berger, 2014; Driver, 1984). As proposed here and in line with the requirements of a statistical investigation, *meta-questioning* involves (a) identifying the contextual issue to be investigated and posing initial questions, (b) interpreting and reviewing/critiquing initial questions, and (c) anticipating data collection and refining questions (as illustrated in Fig. 1). We report on the initial components of an activity in which 4th-grade students engaged in meta-questioning as they created and refined survey questions with the aim of comparing life in two Australian cities. The activity was conducted as part of a four-year longitudinal study that is developing 3rd–6th graders' competencies in modelling with data. Specifically, we address the following:

1. What was the nature of the students' initial posed questions and their subsequent refined questions? How did students justify the changes they made?
2. What forms of required data did the students anticipate for their refined questions?
3. What improvements occurred across the students' successive question refinements?

2017. In xxxx (Eds.). 40 years on: We are still learning! *(Proceedings of the 40th annual conference of the Mathematics Education Research Group of Australasia)* pp. xxxx. Melbourne: MERGA.

Meta-questioning in Statistical Investigations

Statistical investigations are typically described as a four-step process involving posing a question, collecting data, analysing data, and forming and conveying conclusions (e.g., English & Watson, 2015; Konold & Higgins, 2003; Lavigne & Lajoie, 2007; Pfannkuch & Wild, 2003). Typically, such investigations usually do not proceed in such a linear, orderly fashion. As Konold and Higgins (2003) indicated, these phases of a statistical inquiry are interdependent, with a good deal of "backtracking" often taking place (p. 194). That is, experienced researchers consider ahead of time the data they will need, the analyses that might be undertaken, and possible findings. Further refining of initial questions, deciding on data collection and analysis, and how findings will be communicated follow. There is also looking backwards to the initial question/s as data are analysed and findings produced. In essence, data investigations are complex and require early and continued development of statistical thinking skills. Simply treating an investigation in the primary school as a teacher directed, lock-step process does not adequately develop young students' statistical skills and understandings. In particular, the important processes of meta-questioning, which are often ignored or at best completed by the teacher, require substantially more attention. The words of Hanner, James, and Rohlfing (2002) are particularly germane to this issue:

> Very often, teachers solve all the interesting issues for kids and present them already resolved to children, without giving children the opportunity to grapple with such questions as, "What attributes should we include?", "How many attributes should we consider?" and, "How should they be represented?" When teachers take over these decisions, all that's left is a cut-and-dried graphing or sorting activity, in which teachers have done all the intriguing and motivating thinking ahead of time (p. 106).

As highlighted in earlier studies (Konold & Higgins, 2003; Lehrer & Romberg, 1996), primary school students can learn a good deal about data as they tackle issues that arise in creating statistical questions, in particular when they anticipate implementing surveys with questions they create. By encouraging children to think about how they would answer their posed questions, they become aware of the different responses that might be elicited, and the multiple ways in which their questions could be interpreted, especially if they are worded vaguely or ambiguously. Given the need for more research on statistical question posing and refining, especially in the primary school (Allmond & Makar, 2012; Arnold, 2008; Konold & Higgins, 2003) we argue that special attention needs to be given to the processes of meta-questioning (Fig. 1).

Previous research has highlighted the importance of students being able to distinguish between a question that "anticipates a deterministic answer" and one that "anticipates an answer based on data that vary" (Franklin & Garfield, 2006, p. 350). Further emphasising this issue, Konold and Higgins (2003) cite research on the individual and aggregate views in questions students ask. Many of their questions simply require identifying individual perspectives, usually themselves, such as "Who is the oldest?" Konold and Higgins argue that it is not until students are required to respond to questions pertaining to possible differences between two groups that they begin to think about group characteristics rather than those of the individual. To this end, the present activity required students to address the broad contextual issue of comparing life in City A with life in City B.

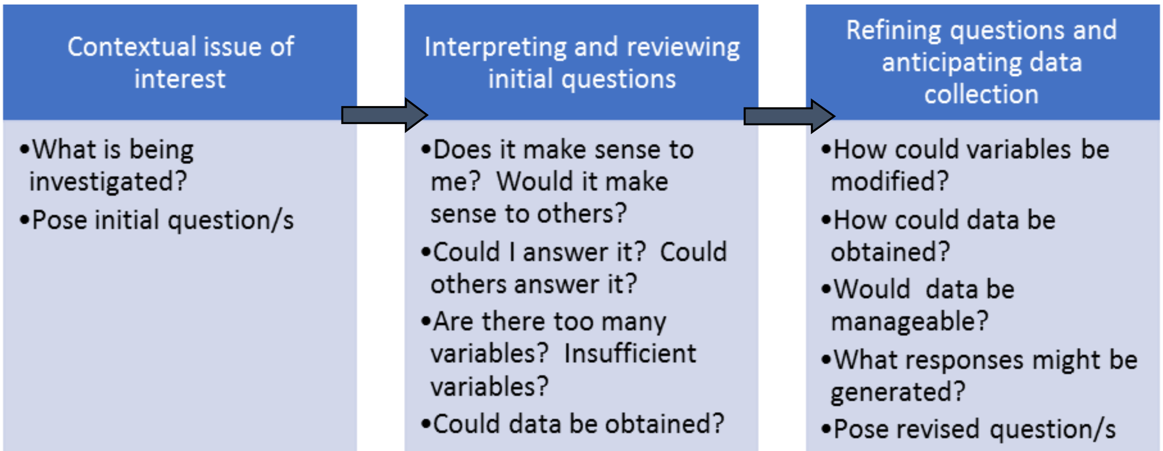

Figure 1: Meta-questioning in statistical investigations

Method

Participants

Four 4th-grade classes and their teachers from two, middle class schools across two Australian states participated in the activity. The school in City A had 56 consenting students (28 in each class) with an average age of 9.8 yrs. The school in city B comprised 33 consenting students (17 in one class, 16 in the other) with an average age of 9.6 yrs.

Research Design

The study adopted a design-based approach involving three main phases, namely, (a) development of activities in consultation with teachers, followed by professional development sessions; (b) teaching intervention implemented by the classroom teachers, together with data collection by researchers, and (c) analyses of data leading to suggestions for future teaching interventions (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003).

Activity implementation. The overall activity was designed to develop students' skills in (a) posing and refining statistical questions in developing a survey to learn more about their peers in each city (specifically, to compare their respective city lives), (b) collecting, analysing, and representing the data from the survey, (c) identifying variation and trends in the data, and (d) drawing conclusions and inferences while acknowledging uncertainty.

The activity comprised four parts. In the first part of the activity, the students "introduced" one another by sharing photos and providing their inter-state peers with some information, for example, about their family life, hobbies, pets, and sports. Two animal "mascots" were chosen to represent the respective states, with class time spent exploring their features and habitats. In the second part, students studied the feedback from their peers, together with viewing video clips displaying features of each city. In the third part, which is reported here, each student posed three questions that would yield information about life in City A compared with that in City B. The students recorded their questions in the workbooks supplied for each student. Students chose one of their three questions to share with their group (comprising 2-3 students) and were reminded to check whether their question would "help us to compare life in City A with life in City B." The notion of *variation* was revisited with respect to the purpose of the questions, with students having explored the notion in the previous year's activities.

Students then refined their questions, indicating how their changes improved their original questions. The students were to reflect critically on their improved question and then further refine it. All responses were recorded in the students' workbooks. Part 3 ended with students recording the data they would need in order to answer their refined question. The remaining part of the activity (not reported here) involved the researchers selecting 22 of the students' posed questions (making a few modifications where necessary) for inclusion in an all-class survey. All classes subsequently completed the survey online, followed by students analysing and representing the data, and drawing conclusions and inferences.

Data Sources and Analysis

Data sources included students' responses in their individual workbooks. All whole class discussions were video-taped, while "focus" groups were video- and audio-taped. Focus groups comprising 2-4 children of mixed achievement levels were chosen in consultation with the teachers. Students' written workbook responses are the primary data source reported here, however, transcripts of focus group work and whole class discussions were analysed in conjunction with these responses.

Content analysis (Patton, 2002) was used by both the first author and an experienced research assistant in initially identifying, coding, and categorising the data recorded in the students' workbooks. In collaboration with the second author and another experienced research assistant, a further cycle of refined coding was undertaken to ensure meaningfulness and accuracy. Transcribed video- and audio-taped data were also analysed to gain greater insights into how students posed, critically analysed, and refined their questions.

In addressing the research questions, we report on (a) the nature of the original questions students chose to refine and their one or two subsequent refined questions, (b) their explanations/justifications for how they refined their original question, and (c) the data that students anticipated would be needed to answer their refined questions.

Results

RQ1: What was the nature of the students' initial posed questions and their subsequent refined questions? How did students justify their improvements to their original questions?

In analysing the students' responses, we considered the features of their initial posed questions with a focus on the question they chose to refine, how they refined their chosen question, their justifications for their changes, and their further (final) refined question. If students posed more than one question within their initial question/s, we recorded only one code for the more appropriate question. The final coding scheme we adopted for both the students' *chosen question to refine* and their *refined questions* follows. Examples and percentages of student responses for each category code are presented.

0*: No response or the content in the original question was subsequently ignored in a refined question.* Sixteen percent of students (N=82) either did not indicate which of their original questions they refined or wrote a refined question that was not related to the original question chosen. All students, however, offered an acceptable first refined question. For the second refinement of the original question—in City A this refinement was optional—45% of students did not provide a response.

1: *The question would generate a simple yes/no response*. Examples included "Does City B have the same weather then [sic] State A?"; "Do you go to theme parks or hotels or go swimming often?" Twenty-two percent of students (N=82) gave a response of this nature for their chosen original question. In contrast, only one student posed such a question for their first refined question and only two for the second refined question.

2*: The question would generate a single numerical or categorical response*. A question of this form appeared superior to the previous type but nevertheless would still produce rather limited data. Such responses included questions of the nature, "How many holidays do you get?"; "What is your favourite food?"; "What is the best thing to do in City B?" Sixteen percent of students posed an original question of this format, with the number increasing to 32% for the first refined question. In contrast, only 7% of students posed a second refined question of this type.

3: *Open-ended questions* that would either (a) generate *multiple responses/lists* (e.g., "What are the typical foods you eat every day?"; "What do you do when it's sunny and what do you do when it's rainy?", and/or (b) ask for an *explanation* (e.g., "Why does City B have theme parks and City A doesn't?"). This third category, which we consider a more advanced form of questioning than the previous types, characterised 41% of the originally posed questions (N=82). This increased to 52% for first refined question, but declined to 39% for the second refined question.

4: *Question gives variables/options to be considered*. Some of the questions coded in this category also included a yes/no query. Sample responses included: "What do you do after you go home after school? Do you do homework or do you play?"; "What is City B's weather like? Is it warm or cold?" This question type was considered the most sophisticated in generating manageable, more substantial data. Only 5% of students (N=82) posed an original (chosen) question of this nature. This improved somewhat for their first refined question, with 15% of students posing such a question. For the second refined question, however, only 6% of the students offered such a response.

In analysing the students' *justifications for their refinements*, we adopted the following categories of codes:

0: *No response or an irrelevant/idiosyncratic response*. Unfortunately, 34% of students could not justify the improvements they made to their original question.

1: *Repeating of question, or the student was satisfied with the posed question, or couldn't refine it.* Eleven percent of responses were assigned to this category.

2: *Recognising the need to change the type of question*. Justifications here included improving on a question that simply requested a "yes" or "no" answer, for example, "My old queschon (question) was a yes or no queschon [question]." Only 2 students (2%) gave this justification.

3: *Changing qualifying terms* or *removing a simple item of information*. Such responses included: "Well I put the words 'or', 'popliar' [popular] in my question because it might not be your favriout [favourite] but some other people might like it more"; "By saying why instead of would"; and "I deleted different in my question and added are they fun celibrations (celebrations)." Six percent of students gave such justifications.

4: *Adding response options,* such as: "It changed my question because i [I] gave them some options; more choices what they can say to us." Only 2 students (2%) offered explanations of this type.

5: *Facilitating data collection through (a) requesting more detail from the responder,* and/or *(b) asking for an explanation*, and/or (c) *clarifying the question to get more information or more precise information*, and/or (d) *recognising that the previous question was irrelevant* for the purpose of comparing life in the two cities. Responses of this type included: "I put some more information so that I could get some more data!"; "I would get a variety of answers which is better because you get lots of different opinions."; "There's more than 1 ansewer [answer]"; and "I changed my first question because they were all erelevent [irrelevant] and they wouldn't collect any data from the set question." Forty percent of students gave justifications of this form.

6: *Recognising the need to delimit variables*. Only four students (5%) offered a justification of this form, for example: "It makes it less broad and more speific [specific]"; "Because now there won't be a lot of data"; and "They improved my question because I added a few words to the first question and I got rid of my second question."

RQ2: What forms of data did the students anticipate for their refined questions?

Only students in City B (N=33) responded to the issue of the data that would be required to answer their refined questions. We coded the students' responses as follows.

1: *Reference to specific data on the question variables*. Responses included: (a) the *type of data*, such as, "The data I would like is temperatures (Numbers e.g 11° to 13°) during the year (winter, sumer [summer], Autumn, Spring), (b) *lists of data*, for example, "The data I would need is a list of celibrations [celebrations] they have and an answer if they are fun and how they enjoyed it", or (c) *an ordering of data*, for example, "The data I need is the most popular sport to the least popular out of the 5. I also want to know the popular sport cultures." Anticipated data of this nature was given by 42% of the 33 students.

2: Reference to the *contextual issue* (i.e., comparing life in the two cities). Students' responses included: "I need exclusive information relating too [to] their lives and I can see if life in City A is the same as life in City B or completely diffrent [different]"; "I need the data to my question to know if the population in the city is bigger then [than] City B or less." Only 9% of students indicated they would need data addressing the contextual issue.

3. Students offered examples of *possible responses to their questions*, such as: "I want them to say two sports/activities they play in winter and why? For example, two activities that they would only play in winter. Maby [Maybe] because it is to (too) hot in sumer [summer] to play the two games they would play in winter." This response was also popular, with 42% of students responding in this manner.

4. Reference to the *nature of questions asked* and *data required* in a survey, such as: "For my question I don't want yes or no, I am aiming to get a [an] explanation. For a good survey, you need a detailed answer. You can also have some varition (variation); Detailed answers, reasons, variation and truth/fact, what type of landscape they live in." Only 2 (6%) of the 33 students offered a sophisticated response of this nature.

RQ3: What improvements occurred across the students' question refinements?

In addressing the third research question we considered the relative changes (ignoring the actual categories of responses) in the students' questions as they progressed from their chosen question, to their first refined question, and then to their second (final) refined question. Only 45 out of 82 students posed three questions (including the original) that could be coded, while 37 students only recorded two questions (original and first refined question). Of the 45 students who recorded three questions, only two did not change their original question and 13 changed once. For the 37 students who only wrote two questions, 34 changed their question.

In coding the relative changes, we considered whether the questions improved, remained the same, declined, or were a combination of these. For those students who offered three questions (N=45), 53% remained in the same coded category across all questions and 33% showed improvement on the third question or at least improvement on the second question. Fourteen percent remained in the same category or declined as they moved towards their third question (i.e., second refined question). For the 37 students who did not record a third codable question, 59% improved, 32% posed questions in the same category, and 8% posed a question of a lesser quality. Finally, across all students (N=82) for their original and first refined questions, 43% improved, 50% remained the same, and 7% declined.

Discussion and Concluding Points

Meta-questioning is a significant, underrepresented component of statistical investigations, which needs to be introduced early and developed across the primary years. We have reported on 4^{th}-grade students' meta-questioning as they created and refined survey questions for comparing life across classes in two Australian states. Of interest were the nature of the students' initial posed questions and their subsequent refined questions, students' justifications for their refinements, their anticipated data collection, and improvements in their question posing as they subsequently refined their questions.

In coding the form of questions students posed, we identified a hierarchy of increasingly sophisticated question types, ranging from a question that would generate a simple yes/no response to one requiring a single numerical or categorical response, through to open-ended questions, and finally questions that offer a set of variables or options to be considered. As students progressed from their original posed question to their first and second refined questions, the use of simple "yes"/no" questions declined considerably. Students' responses suggested that they appreciated that the data generated would be rather limited. Requesting a single numerical or categorical response was more prevalent in the first refined questions, but again the data indicate that the students are capable of rethinking their questions and offering alternatives and feasible justifications. The most common question type was one that was open-ended, where a range of responses would potentially result in a large collection of difficult-to-manage data. The most sophisticated question type, where a choice of variables is provided, was created by only a few students.

Overall, the findings reveal the difficulties primary school students can experience in creating statistical questions, but with opportunities to engage in meta-questioning, as conceptualised in this paper, they can develop key understandings that are foundational to productive investigations. Students need experiences in generating their own questions rather than being supplied with those created by others. Furthermore, when students engage in contextual issues that involve questions regarding potential differences between two groups, they can learn to distinguish between group characteristics that vary and individual perspectives that produce deterministic answers (Franklin & Garfield, 2006; Konold & Higgins, 2003).

Although we did not report on the remaining activity components, we argue that meta-questioning should occur throughout an investigation. In exploring issues of interest, students should reflect on the question/s they are investigating as they collect and analyse data, and consider the conclusions and inferences that can be drawn. Teacher support is important in fostering students' skills and dispositions to engage in meta-questioning, as reflected in earlier studies (e.g., Allmond & Makar, 2010; Lehrer & Schauble, 2002). Clearly, more opportunities are needed for younger students to progress from posing questions that would yield either limited data or too much data, to ones that would generate substantial yet manageable data. The hierarchy of question types we have identified provides one framework for analysing and developing core investigative questioning skills, which remain largely under-developed and under-researched in many primary classrooms.

Acknowledgements

The present study was supported by an Australian Research Discovery Project grant #150100120. The opinions expressed in this report are those of the authors and not the Council. We acknowledge the contributions of Jo Macri and Suzie Wright in assisting with data coding, as well as the participation by the students and classroom teachers.

References

Allmond, S., & Makar, K. (2010). Developing primary students' ability to pose questions in statistical investigations. In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. In *Proceedings of the 8th International Conference on the Teaching of Statistics*, Ljubljana, Slovenia, July 11-16). Voorburg, The Netherlands: International Statistical Institute.

Arnold, P. (2008). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. *Proceedings of the 11th International Congress of Mathematics Education*, Monterrey, Mexico, July 6-13, 2008. Retrieved from http://tsg.icme11.org/tsg/show/15

Berger, W. (2014). *A more beautiful question*. New York, NY: Bloomsbury.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9-13.

Driver, J. L. (1984). Metaquestions. *Noûs*, *18*(2), 299-307.

English, L. D. (2014). Statistics at play. *Teaching Children Mathematics*, *21*(1), ~~pp.~~ 37-44.

English, L. D., & Watson, J. M. (2015). Statistical literacy in the elementary school: Opportunities for problem posing. In F. Singer, N. Ellerton, & J. Cai (Eds.), *Problem posing: From research to effective practice* (pp. 241-256). Dordrecht: Springer.

Franklin, C. A., & Garfield, J. B. (2006). The GAISE Project. Developing statistics education guidelines for grades pre-K-12 and college courses. In G. F Burrill & P. C. Elliot (Eds.), *Thinking and reasoning with data and chance* (pp. 345-375). Reston, VA: National Council of Teachers of Mathematics.

Hanner, S., James. E., & Rohlfing, M. (2002). Classification models across grades. In R. Lehrer, & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 99-117). New York, NY: Teachers College Press.

Lavigne, N. C., & Lajoie, S. P. (2007). Statistical reasoning of middle school children engaged in survey inquiry. *Contemporary Educational Psychology, 32*, 630-666.

Lehrer, R., & English, L. D. (in press). Introducing children to modeling variability. In D. Ben-Zvi, J. Garfield, & K. Makar, (Eds.), *International handbook of research in statistics education*. Dordrecht, The Netherlands: Springer.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition & Instruction*, *14*(1), 69-108.

Lehrer, R., & Schauble, L. (2002). (Eds.). *Investigating real data in the classroom: Expanding children's understanding of math and science.* New York, NY: Teachers College Press.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Patton, M. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Pfannkuch M., & Wild, C. (2003). Statistical thinking: How can we develop it? *Proceedings of the 54th International Statistical Institute Conference*, Berlin, Germany, 13-20 August, 2003. [CDROM]. Voorburg, The Netherlands: ISI. Retrieved from www.stat.auckland.ac.nz/~iase/.publications/3/3235.pdf

Watson, J., & English, L. D. (2015). Introducing the practice of statistics: Are we environmentally friendly? *Mathematics Education Research Journal*, *27*(4), 585-613. DOI 10.1007/s13394-015-0153-z

Whitin, D. J., & Whitin, P. E. (2011). *Learning to read numbers: Integrating critical literacy and critical numeracy in K-8 classrooms.* New York, NY: Routledge.