CRITIQUING INVESTIGATIVE QUESTIONS

Pip Arnold[1] and Maxine Pfannkuch[2]
[1]Cognition Education Limited, New Zealand
[2]University of Auckland, New Zealand
parnold@cognitioneducation.com

A "good" investigative question is one that allows for rich exploration of the data in hand, discovery, and thinking statistically. Developing and using the criteria for what makes a good investigative question was one of the outcomes of research in a year 10 (age 14–15) class. Using these investigative question criteria, this paper focuses on students' ability to critique investigative questions posed by others. Pre- and post-test responses of 27 14-year-old students are explored and the quality of students' critique is discussed, including the implications for supporting their ability to pose their own investigative questions.

INTRODUCTION

The role that posing investigative questions plays in assessment for qualifications in New Zealand and the identification of a lack of teacher knowledge in this area (Arnold, 2008) had highlighted posing investigative questions as a problematic situation. Teachers need to know the components and concepts underpinning a good investigative question, and the learning that students need to be immersed in to support their posing good investigative questions. A "good" investigative question is one that allows for rich exploration of the data in hand, discovery, and thinking statistically. A good investigative question allows students to engage in interesting work and has an element of open-endedness. Over three research cycles what makes a good investigative question was explored, including different ways to introduce students to the elements that make up a good investigative question. This paper focuses on the last of these research cycles and explores the research question – What level of critique are students in year 10 (ages 14–15) making on investigative questions posed by others?

LITERATURE REVIEW

In this brief literature review the nature of investigative questions are discussed followed by some references to studies where students critiqued mathematics examples.

Investigative questions

In the big picture of statistical enquiry, the investigative question is the statistical question or problem that needs answering or solving. The process of developing or creating the investigative question is iterative and requires considerable work to get it right (e.g., delMas, 2004; Franklin et al., 2005; Wild & Pfannkuch, 1999). Posing investigative questions has been identified as a problem area for students. To get precise investigative questions that can be correctly interpreted and that yield useful information, an *interrogative* process, which involves asking questions of the investigative question, is necessary (e.g., Burgess, 2007; Graham, 2006; Konold & Higgins, 2003). For example, Burgess (2007) acknowledges that some of the specialized content knowledge teachers need relates to their ability to be able to decide if a question posed by their students is suitable, unsuitable, or whether changes can be made to make the question suitable. He notes that teachers need to ask whether the students will find the investigative question interesting. Graham (2006) provides five useful considerations for forming a good investigative question. These considerations pick up several different aspects of interrogating the investigative question. The considerations are whether the question is: "(1) actually a question, rather than simply an area for investigation…; (2) personally interesting to you…; (3) likely to draw on data that will be available within the time frame of the investigation…; (4) specific, so that it is answerable from data…; (5) measurable…." (p. 88).

Investigative questions differ from survey or data collection questions, though both investigative and survey questions are questions that are formally posed with a specific purpose in mind. Investigative questions are questions that are asked of the data, whereas survey questions are questions that are asked to get the data.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Using examples, critique and correction

Previous studies have explored the use of erroneous examples to facilitate learning of mathematics (e.g., Durkin & Rittle-Johnson, 2012; Siegler, 2002) and have found that it can make a difference. For example, Durkin and Rittle-Johnson (2012) found that students who had worked with correct and erroneous examples were more likely to discuss correct concepts than those students who had worked with correct examples only.

Evans and Swan (2016) report on a study where students were asked to interpret, complete, compare, and critique pre-prepared, hand written "sample student responses" to non-routine unstructured problems. They found that students could draw out affordances and limitations of each but did not necessarily decide which one was correct. They found that many students did critique the suitability of the sample response within the context of the problem.

METHOD

The research method followed design research principles (Roth, 2005) for a teaching experiment in a classroom. In the preparation and design stage the first author developed the teaching and learning materials for a 16-lesson teaching experiment in conjunction with the classroom teacher, and considered relevant literature. Both the classroom teacher and first author were involved in the implementation of the activities in the teacher's year 10 (ages 14–15) class. Following each lesson there was reflective discussion and adjustments were made as needed to the learning trajectory. The learning activities were designed to support students' understanding of posing statistical investigative questions and built on research work previously undertaken (Arnold, 2008; Burgess, 2007). In order to focus students' attention on developing criteria for what makes a good investigative question, a class activity was developed in which students critiqued investigative questions posed by others.

Posing investigative questions was the specific focus of lesson 5. The learning trajectory was specifically designed to support students to develop the following criteria for what makes a good investigative question: (C1) the variable(s) of interest is/are clear and available, (C2) the population (or group) of interest is clear, (C3) the intent is clear, (C4) the question can be answered with the data, (C5) the question is one that is worth investigating, it has a purpose, and (C6) the question allows for analysis to be made of the whole group (Arnold, 2013). Note that these criteria were developed in the previous two research cycles.

The approach to teaching posing investigative questions in the third research cycle was one where the teacher gave the students questions that had been posed by others before she required them to pose their own. In lesson 5 the students had to sort several investigative questions into groups providing a catalyst to talk about what questions were good questions and what questions were not. From this discussion, some of the criteria that had previously been established by the research were re-established by the students. That is, the students and teacher developed the criteria based on the class discussion about the questions they were sorting. Criteria that the students came up with included that the question needed to be about the overall distribution of the data (Criterion 6: C6), it must be interesting (C5), and the variable (C1) and group needed to be stated (C2). Student reflection at the end of the lesson elicited a further criterion that had not been mentioned in class: that the type of question needed to be clear (C3). At this point the teacher resisted the urge to "finish" the criteria and settled to leaving the remaining criterion (C4) until it naturally arose in the teaching and learning sequence (Arnold, 2013). The teacher used the criterion to reflect on with the students as to whether a suggested investigative question would be suitable or not as the unit of work progressed.

The 29 students in the class were above average in ability and from a mid-size (1300), multicultural, mid socio-economic inner-city girls' secondary school. Before and after the teaching experiment the students were given a pre- and post-test, one item of which was to critique five investigative questions that had been posed by others. There were four "bad" questions (see Table 2) and one "good" question. The instructions were: *The following questions were posed about the data set given in Table [Table showed a multivariate data set with eight variables and data for 24 cases from a sample of 254]. For each question comment on whether you think the question is a good question or not. Give reasons for why or why not. If the question is not a good question, change it to make a better question*.

Students' pre-and post-test responses to the five questions were graded using the SOLO taxonomy (Biggs & Collis, 1992; Watson, 2005). The descriptors for student critiquing responses were developed through a process of moving between the criteria for what makes a good investigative question and student responses. Briefly, the descriptors with numerical score for grading the student responses were: **no response** (NR-0); **pre-structural** (PS-1) – attempts to answer the investigative question OR tries to pose, incorrectly, a new investigative question; **uni-structural** (US-2) – gives one correct piece of evidence, either a feature correctly identified or that the investigative question was a good question or a bad question; **multi-structural** (MS-3) – gives two pieces of evidence OR relational evidence with an incorrect population; **relational** (R-4) – gives three pieces of evidence, including at least two of: features identified, good or bad question or improved a bad question correctly; **extended abstract** (EA-5) – gives at least two features, identified good or bad questions correctly and has an improved question that is at the top level of questions.

RETROSPECTIVE ANALYSIS

All students' pre- and post-test responses to the critiquing questions item were analyzed for this paper. Each of the five questions were individually graded using the SOLO taxonomy and an overall mean for the pre-test responses and an overall mean for the post-test responses was calculated for each student. Figure 1 shows a scatter plot of the average pre-test score (AveragePre) versus the average post-test score (AveragePost) for each student. The line (AveragePost=AveragePre; y=x) represents where the two averages are the same. All 27 students are on the upper left side of the y=x line signaling that all students' average score improved from their pre-test to their post-test. Note in the scatter plot there are 24 dots, due to three situations where the scores have doubled up (AveragePre,AveragePost; 2,2.8; 2.17,3.4; 2.17,4.2).

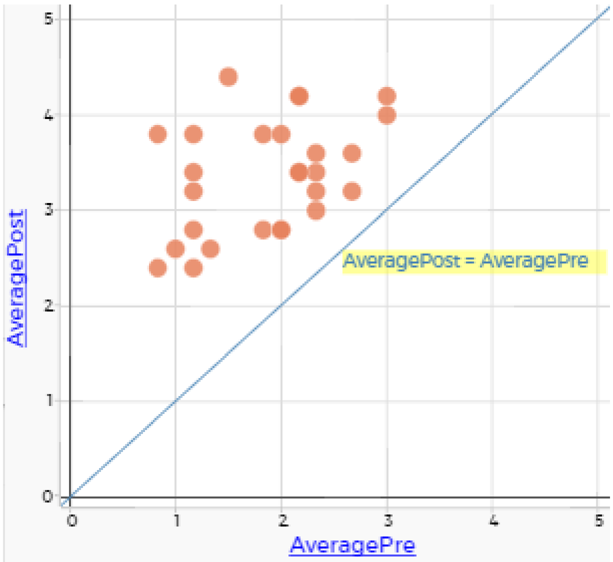

Figure1: Student pre- and post-test average scores across five questions

Each question was also individually analyzed to look at the difference in SOLO grades from pre-test to post-test. Table 1 gives a summary of the differences from pre-test to post-test for all five questions. Table 1 shows that on average there was improvement in all five questions from pre-test to post-test. For question 1 and 2, 89% of the students (n=24) improved their SOLO grade by at least one from pre-test to post-test. For question 3 it was 92% (n=25) improvement, for question 4 it was 74% (n=20) and for question 5 it was 52% (n=14). Question 5 stands out in that the question posed was not able to be answered by the data available, a fact that many students picked up in the pre-test, hence the high pre-test SOLO mean (2.22). However, they did not improve a lot in the post-test, with most students still picking up the obvious error, but not adding anything further, for example noticing that the population was not clear.

Table 1: Differences between pre-test and post-test for individual questions

Question	Difference in SOLO score from pre-test to post-test for individual students							Mean SOLO pre-test	Mean SOLO post-test	Difference of means
	-2	-1	0	1	2	3	4			
1			3	3	12	8	1	1.59	3.63	2.04
2			3	8	12	4		2.07	3.7	1.63
3	1		1	9	7	8	1	1.63	3.44	1.81
4		2	5	11	5	4		2.04	3.19	1.15
5	2	1	10	10	2		2	2.22	2.85	0.63

Table 2 gives examples of student responses to the four "bad" questions in the pre- and post-tests, with commentary on what evidence is present and the difference in SOLO grades from pre- to post-test.

Table 2: Examples of student responses to the four "bad" questions

Question	Response (SOLO grade in brackets)	Commentary
1.(Pre) Do girls have longer popliteal lengths than boys?	Student A: Typical neck circumference girl - 29cm. Typical neck circumference boy - 45cm. Typical neck circumference students - 34 cm (PS-1) Student B: No, boys do (PS-1)	Both students attempted to answer the investigative question given.
1.(Post) Do girls have longer armspans than boys?	Student A: This investigative question is not good because a) it does not specify which population of girls/boys (e.g. their year level/country) e.g. I wonder if the typical year 10 NZ girls arm span tend to be longer than the typical arm span of yr 10 NZ boys? (R-4) Student B: This isn't. It doesn't show the population and is a very free question. I wonder if year 10 NZ girls tend to have longer arm spans than year 10 boys? (R-4)	Both students identified that it is not a good question, that the population is not specified (feature) and improved the question. Both these students improved from a SOLO grade 1 to a SOLO grade 4. The difference is 3.
2.(Pre) Who has the biggest armspan?	Student C: This isn't a good investigative question because you haven't been given enough info on who the person is. It should be on average which gender has the biggest arm span? (MS-3)	This student identified that it is not a good question and has implied that the question should include more than the individual.
2. (Post) Who has the longest popliteal length?	Student C: This isn't a good question because it's asking for a "who" and that isn't looking at the overall data. It also doesn't give the population. It should be: What are the typical popliteal lengths of yr9-13 NZ girls? (EA-5)	The student has identified that it is not a good question. She specified that it should be about the overall data and that the population is missing (two features). She improved the question. The difference is 2.
3.(Pre) What are typical neck circumferences for these students?	Student B: 30cm, 33cm, 35cm (PS-1) Student C: ? (NR-0)	Student B attempted to answer the investigative question given, Student C made no attempt.

3.(Post) What are typical neck circumferences for these students?	Student B: Doesn't say WHAT students. What are typical neck circumferences for year 10 NZ students? (R-4) Student C: This isn't good because it doesn't give the population. It should be: What are the typical neck circumferences for yr 9-13 NZ boys? (R-4)	Both these students identified that it is not a good question, identified one feature is missing, the population, and improved the question. The difference is 3 and 4 respectively.
5.(Pre) What is the most popular sport played?	Student D: No this isn't a good question because the data sheet doesn't contain any info about sports being played. You could change it to something about fitness levels. Eg. In which region are fitness levels the highest (MS-3)	The student identified that it is not a good question, and identified the data is missing (a feature). Her improved question is not suitable.
5.(Post) What is the most popular sport played?	Student D: No because the data sheet doesn't even collect data for what sport each student plays, doesn't have the population or a correct variable. A suitable question related to sport could be: what are typical pulse rates for yr 9-13 NZ students? (EA-5)	The student identified that it is not a good question, identified two features – population missing and data missing (correct variable), and improved the question. The difference is 2.

In the pre-test students were critiquing investigative questions posed by others mostly at the pre-structural or uni-structural level. This is indicated by the mean pre-test score for each question (see Table 1) and the average pre-test scores for the students. Twenty-five of the students had an average pre-test score less than 3 (an average of 3 indicates working at the multi-structural level). For the 27 students, across the five questions there were 35 responses that scored a 3 – multi-structural thinking – and three responses that scored a 4 – relational thinking. This is a total of 38 responses at the multi-structural level or above out of a possible 135 responses (28%).

By the post-test all questions showed an improvement in the level of critique. Question one showed the biggest mean improvement and question five showed the smallest mean improvement (see Table 1). Generally, by the post-test students were working at a multi-structural or relational thinking level. For the 27 students, across the five questions in the post-test there were 46 responses that scored a 3 (MS), 53 responses that scored a 4 (R) and eight responses that scored a 5 – extended abstract thinking. This is a total of 107 responses at the multi-structural level or above out of a possible 135 responses (79%).

The graph in Figure 1 shows that 19/27 (70%) students were on average working at the multi-structural level or above by the post-test compared with 2/27 (7%) students in the pre-test.

CONCLUSION

Students ages 14–15 can identify if an investigative question is a "good" or "bad" investigative question. They can name at least one feature missing for "bad" questions or at least one feature present for "good" questions. The level of critique that these students in year 10 (ages 14–15) make on investigative questions posed by others is at least at the multi-structural level (SOLO taxonomy), with two-thirds of them reaching this level by the post-test. Nearly 20% of the students are providing critique at the relational thinking level in the post-test, indicating that in addition to identifying a "good" or "bad" question and at least one feature they have either improved the investigative question or identified another feature. Furthermore, these students were asking questions of the investigative questions, a reasoning process that Burgess (2007) noted was often lacking in pre-service teachers. Overall these students showed a reasonable depth of understanding of the criteria of what makes a good investigative question, which suggests that critiquing questions may indirectly be facilitating their learning about what makes a good investigative question (cf. Durkin & Rittle-Johnson, 2012).

A good investigative question allows students to engage in interesting work and has an element of open-endedness, it also allows for rich exploration of the data in hand, discovery and thinking statistically. Indeed, Gould, Bargagliotti and Johnson (2017) suggest that developing questioning skills in students, including the crafting of productive investigative questions, is critical for success in analyzing, interpreting and drawing conclusions from data. We conjecture that critiquing investigative questions posed by others may enhance student understanding of the features or criteria of what makes a good investigative question and that a critiquing approach shows promise for introducing posing investigative questions to students. Critique also supports students as they start to think like a statistician, interrogating different elements of the statistical enquiry cycle, as they work through a statistical investigation.

REFERENCES

Arnold, P. (2008). *What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation.* Paper presented at the 11th International Congress on Mathematical Education (ICME-11), Monterrey, Mexico. http://tsg.icme11.org/document/get/481

Arnold, P. (2013). *Statistical investigative questions: An enquiry into posing and answering investigative questions from exisiting data.* (Doctoral thesis.) The University of Auckland, New Zealand.

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

Burgess, T. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* (Doctoral thesis.) Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/dissertations/07.Burgess.Dissertation.pdf

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 79–95). Dordrecht, The Netherlands: Kluwer.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206–214.

Evans, S., & Swan, M. (2015). Developing student questioning when problem solving: The role of sample student responses. In K. Krainer & N. Vondrova (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3015-3021). Prague, Czech Republic: Charles University in Prague, faculty of Education and ERME.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association.

Graham, A. (2006). *Developing thinking in statistics*. London, England: Paul Chapman.

Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers' reasoning with participatory sensing data. *Statistics Education Research Journal, 16*(2).

Konold, C., & Higgins, T. (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.

Roth, W.-M. (2005). *Doing qualitative research: Praxis of method*. Rotterdam, The Netherlands: Sense.

Siegler, R. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.). *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge, NY: Cambridge University.

Watson, J. (2005). Developing an awareness of distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–33). Brisbane, Australia: University of Queensland.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265. doi: 10.1111/j.1751-5823.1999.tb00442.x