INFORMAL INFERENTIAL REASONING

Maxine Pfannkuch
The University of Auckland, New Zealand
m.pfannkuch@auckland.ac.nz

Year 11 (15-year-old) students are not exposed to formal statistical inferential methods. Therefore, when drawing conclusions from data, their reasoning must be based mainly on looking at graph representations. This study investigates the type of reasoning that might develop students' informal inferential statistical reasoning towards a more formal level. A perspectives model is developed for a teacher's informal inferential reasoning from the comparison of boxplots. The model is then used to analyse her students' responses to an assessment task. The resultant analysis produced a conjectured hierarchical model for students' reasoning. The implications of the findings for instruction are discussed.

INTRODUCTION

Informal inferential reasoning is interconnected to reasoning from distributions, reasoning with measures of centre, and sampling reasoning within an empirical enquiry cycle. All these aspects are underpinned by a fundamental statistical thinking element, consideration of variation (Pfannkuch, 2005a). The term *informal inference* is used here to describe the drawing of conclusions from data that is based mainly on looking at, comparing, and reasoning from distributions of data. Traditionally, up to Year 11, students learn about the descriptive use of statistics and graphic displays and are not encouraged to draw conclusions when comparing data sets. At Year 13, students are introduced to formal methods of inference such as confidence intervals, significance testing, and regression models to enable them to draw conclusions from data. With the emphasis now on exploratory data analysis and students being encouraged to be data detectives and to know the purpose, power, and limits of statistical investigation the drawing of conclusions from data is no longer sacrosanct.

Research is only just grappling with understanding the conceptual building blocks for informal inferential reasoning as a pathway towards formal statistical inference, not only in terms of the discipline, but also in terms of student cognition. Without attention to the conceptual underpinnings of and to the devising of pedagogical experiences towards formal inference, inferential reasoning may continue to be problematic for students. David Pratt (personal communication, 7 July 2005) suggests that informal inference requires students to be aware of the game being played. He believes a large part of students' difficulty in understanding the reasoning being used is that the game being played is not being made explicit to the students by the teacher. The students may believe they are reasoning only about the data under consideration, which Pratt refers to as *game one*, whereas the teacher believes that the data are a sample from a population, which Pratt calls *game two*. It is the playing of *game two* that will lead students towards formal inferential reasoning. Activities such as growing a sample or experiencing sampling variability are now being suggested as ways to improve students' sampling reasoning. Another problem is that statistics education, until now, has shied away from informal inference and has not developed a language with shared meanings, nor a shared understanding of how to talk about graphs. Whether the research is focussing on students' cognition by using innovative technology such as *Fathom* (Key Curriculum Press Technologies, 2000) or using students' own products the problem of communicating and articulating the meaning of the statistical representations remains difficult.

Friel, Curcio, and Bright (2001) consider that research is needed on understanding what it is about the nature of reasoning that makes comparing data sets such a challenging task. Furthermore, they believe that graph comprehension involves an interplay between visual shapes, visual decoding, judgement, and context. In particular, informally drawing conclusions from the comparison of boxplots is conceptually demanding since boxplots condense, summarise, obscure information and incorporate statistical notions such as median and quartiles (Bakker, 2004). Biehler (2004) in his research on students' reasoning with boxplots identifies shift, spread, and summary views. He argues that summary views are inherent in the nature of the boxplot representations and lead students to reason intuitively with the cut-off points. He noted that students showed a lack of a shift view, of comments about spread, and that the notion of the median being representative of the data set was difficult to understand.

As part of a larger project on developing Year 11 students' statistical thinking the following research questions are addressed in this paper. When making informal inferences from the comparison of boxplots:

- What views does a teacher communicate in class to her students?
- What views do students communicate on an assessment task?
- What levels do students perform at on an assessment task?
- What are the connections between the students' and teacher's communication?

RESEARCH METHOD

The research project is a two-year study on developing Year 11 students' statistical thinking and is focussed on one teacher and her class. The research method is developmental in that an action-research cycle is set up whereby problematic areas are identified by the teacher, researcher, and students. After implementation of a four-week statistics teaching unit in 2003, half of the students' responses to an open-ended questionnaire, the teacher, and the researcher identified that drawing conclusions from data was problematic. In the 2004 implementation of the statistics-teaching unit, the teacher made a conscious effort to communicate and articulate what she was looking at and how she was thinking when comparing two boxplots as well as writing down the justifications for her conclusion.

The school in which the project is based is a multicultural, secondary girls' school. In the study class of 29 students, 40% were Pakeha (New Zealand European), 30% were Maori or Pasifika, and 30% were Asian or Indian. The teacher considered the students to be average in mathematical ability. In Year 10, students are introduced to the graphing of boxplots. In this class of Year 11 students about 25% of the students were new immigrants, many of whom have English as their second language, and whose previous schooling did not include exposure to boxplots. No technology was available to the students or teacher.

The teacher is Pakeha, in her mid-thirties, and has taught secondary mathematics for twelve years. The class is taught mathematics by the teacher for four hours per week. The teacher is in charge of Year 11 mathematics and therefore, in consultation with the other Year 11 teachers, writes an outline of the content to be covered together with suggested resources and ideas for teaching the unit. She also writes the internal assessment tasks, one of which is statistics, which are moderated at the national level. The researcher previously knew the teacher on a professional basis. The researcher was used as a source of teaching ideas before and during the teaching of the unit and was consulted about the statistics assessment tasks. The researcher videotaped and transcribed the 15 lessons involved in the 2004 statistics-teaching unit. Students' bookwork, assessment responses, and questionnaires were also gathered as data.

This paper focuses on the students' assessment responses to a task involving comparison of boxplots and compares them to the teacher's verbal and written communication. The teacher took the opportunity to confirm or refute the classifications and interpretations of her teaching episodes and the student assessment data. No independent researcher was available, however, to code the student data.

RESULTS

Teacher Communication

From three teaching episodes on the comparison of boxplot distributions, a qualitative analysis of the teacher's communication extracted ten categories for the views adopted. The views adopted by the teacher are briefly summarised in Figure 1 (for a fuller account, see Pfannkuch, 2005b). The eight comparative views for comparing boxplots are non-hierarchical, are interdependent but distinguishable, and are moderated by two other views. This means that the moderator views, the referent and evaluative views, are contained within each of the eight views. For example, for the shift view, "the female graph is slightly higher than the male graph" the referents are "female" and "male" and the evaluative view is expressed by the use of the word "slightly," as the strength of the evidence is assessed. Within some categories there are subcategories, not all of which are illustrated below.

COMPARATIVE VIEWS	
1. Hypothesis	Compares and reasons about the group trend.
2. Summary	Compares equivalent 5-number summary points. Compares non-equivalent 5-number summary points.
3. Shift	Compares one boxplot in relation to the other boxplot and refers to the comparative shift.
4. Spread	Compares and refers to the type of spread/densities locally and globally within and between boxplots.
5. Signal	Compares the overlap of the central 50% of the data.
6. Sampling	Considers sample size, the comparison if another sample was taken, the population on which to make an inference.
7. Explanatory	Understands the context of the data, considers whether the findings make sense, considers alternative explanations for the findings.
8. Individual case	Considers possible outliers, compares individual cases.
MODERATOR VIEWS	
9. Referent	Group label, data measure, statistical measure, data attribution, data plot distribution, contextual and statistical knowledge.
10. Evaluative	Evidence described, assessed on its strength, weighed up.

Figure 1: Teacher's Perspectives Model for Reasoning from the Comparison of Boxplots

The goal of the teacher is to make an inference about a population from a sample through comparing distributions and to justify that inference. Since informal inferences are being drawn from the comparison of two boxplot distributions, visual and feeling cues are used by the teacher. She gradually builds up, in her communication, the multifaceted ways in which she looks at and interprets the comparison of the data sets. For the purposes of making comparisons with the students' views, the signal, sampling, referent, and evaluative views will be briefly elaborated upon.

In the signal view the teacher uses the middle 50% of data as the rough signal amongst the noise. She compares the overlap of the middle 50% of data by drawing double-arrowed lines in both "central boxes." These drawn lines could be conceived of as intuitive visual foundations for confidence intervals for population medians and for significance tests where the differences in centres are compared relative to the variability. The sampling view taken by the teacher is entirely hypothetical whereby she asks the students to imagine what the graphs might look like if another sample of people is chosen, or if the sample size of one plot is similar or much smaller than the other plot. The referent view involves a constant back-and-forth switching between the visual symbol system, the boxplot, and the concepts and ideas to which it refers. For example, other reference systems are the invisible dotplots underneath the boxplots, or the statistical measures such as the median or the data measures. The analysis of the teacher's use of referents raised questions about whether her language sufficiently conveyed the underlying plot. In fact an analysis of her three written conclusions, which the students copied into their books, and which were coded in the same way as the student assessment data, revealed that her referent view was not at Level 3, the highest level.

The evaluative view expressed by the teacher involved much weighing of the evidence before making a decision. In an example where she compared the boxplots of male and female IQ she said: "I've got some conflicting information, the median – females are more clever, but when I look at the whole graph, the whole graph's a bit higher for males … so I'm not ready to say, yes, males have a higher IQ than females." Since the situation appeared to be inconclusive the teacher wrote down: "We are not certain that males have a higher IQ." However, in response to a student who queried this statement, she added: "There is some evidence to suggest that males have a higher IQ for these Uni. students." Conflicting evidence led to a conflicting decision, notably a *game two* decision about the populations, followed by a *game one* decision about the samples.

Student Communication

Part of the student assessment task is illustrated in Figure 2. For this problem the students were also asked to calculate the interquartile ranges for both the Telecom and Vodafone graphs. The students' data were categorized into three levels of comparison for each view, from which emerged four classifications for the overall level of a student. The overall levels are: Level 0: Points Reader; Level 1: Shape Comparison Describer; Level 2: Shape Comparison Decoder; Level 3: Shape Comparison Assessor. All 29 students were able to read appropriate values from the graphs for the interquartile ranges. Hence five students were coded as being at Level 0, a points reader. The other 24 students' level of performance is illustrated in Figure 3 with examples of some responses. Shape describers seem to be reasoning only with the "pictures," entirely ignoring statistical concepts and the plots of the data the boxplots are representing. Shape decoders are beginning to use statistical language to describe the data, to identify features of the plots that could be used as evidence for informal inference, and to ascertain the strength of the evidence. Shape assessors are using statistical language more fluently to describe the data, comparing appropriate features as evidence, and beginning to make judgements on relevant evidence.

There are two main Cell phone companies. They provide text messaging services. Hazel has a Vodafone contract and pays 20 cents for each text message she sends. Kathy has a Telecom contract and she can send up to 500 text messages for a monthly fee of $10.
Loren is interested in comparing the number of text messages sent by Telecom and Vodafone users. She surveyed 100 people (approx. equal numbers of Vodafone and Telecom users). She has displayed the results using box-and-whisker plots.
Loren asked the question: "Which phone company users generally sent more text messages in the last month?" Respond to her question.
How does the graph show this? Make 3 statements to explain.

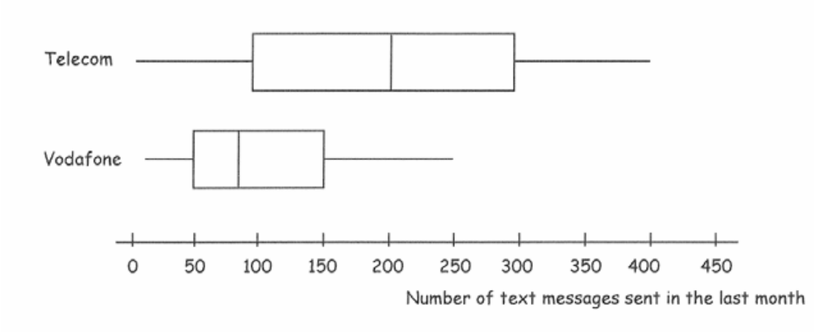

Figure 2: Student Assessment Task

First, the students were coded on one of three levels for each view. Interestingly, ten students were categorised as having a signal view, while seven students used the term "overlap," a view and term not expressed by the previous year's students (Pfannkuch, 2005b). For over half the students their only referents were the group label and statistical measures while one student achieved a Level 3 in this view. For the evaluative view eight students were assessing the strength of the evidence whereas four students attempted to weigh the evidence.

Second, once the levels for each view were decided, criteria were used to decide the overall level of performance. For example, if a student exhibited three different levels over three views then the middle level was deemed to be the overall level. If a student presented only a summary view then the overall level ascribed was one less than the calculated level. Apart from eight students who were entirely on Level 1 for each view expressed, the other students appeared to be on a continuum of understanding about how to communicate their reasoning from boxplots.

Comparison of Teacher and Student Communication

When comparing views of the teacher and students it should be noted that the hypothesis and individual case views are not applicable to the student assessment task. Also, since the assessment question asked the students to explain how the graph showed a particular hypothesis it is difficult to activate the explanatory view. The sampling view was not expressed by any student, and only by the teacher orally, not in her written conclusions. Hence the views expressed by the students resembled the teacher's written views. The referent views of the students and teacher appeared to be similar. The evaluative views of the teacher were strongly present orally and were at Level 3 in the written form, whereas only four students achieved this Level.

View (No. of students expressing view)	*Shape Comparison Describer Level 1*	*Shape Comparison Decoder Level 2*	*Shape Comparison Assessor Level 3*
Summary (26)	Telecom has higher median, UQ, LQ, and max point.	Telecom's median is much higher than the Vodafone median.	Over 75% of T users sent more messages than 50% of V users.
Spread (17)	And the spaces in the Telecom box are bigger than that of Vodafone.	Telecom has a larger IQR. The no. of text messages sent is more spread out.	T's data is more evenly spread out, whereas V's data clusters between the median and LQ.
Shift (4)	Telecom is more to the higher points of the scale.	The whole T graph is higher than V, except for their minimum, which is the same.	Telecom graph is generally higher than the Vodafone graph.
Signal (10)	Telecom box overlaps Vodafone.	The bulk of the data for T is much further along than V.	Central bulk of the data does not overlap that much.
Referent	Refers to names of groups only (Telecom, Vodafone).	Refers to name of data (text messages) and data in 1 view only.	Refers to the data or data plots underneath boxplots in 2+ views.
Evaluative	Compares the difference by description ("higher," "overlaps").	Ascertains strength of evidence ("a lot higher," "much further along").	Ascertains strength of the evidence, and weighs the evidence (not too much overlap).
No. of students attaining level	13	7	4

Figure 3: Conjectured Hierarchical Model for Students' Reasoning from the Comparison of Boxplots with Examples of Students' Responses to Assessment Task

DISCUSSION

Four research questions were posed for this study concerning informal inference from the comparison of boxplots. Each question will be briefly discussed, namely, the teacher's views, the students' views, the conjectured level descriptors for student performance, and the connections between the students' and teacher's communication.

First, the teacher's views reveal that multifaceted reasoning is necessary for articulating the messages contained in the boxplot representations. These views confirm Biehler's (2004) findings but extend his taxonomy beyond the properties of the boxplot distributions to include the hypothesis, sampling, signal, explanatory, referent and evaluative views. Second, the students' views are a subset of the teacher's oral views but are the same views she presented in the written form. Third, the reasoning levels ascribed to the students may result from a combination of cognitive development and method of instruction. The students' performance may reflect both a visual and language developmental pathway, whereby the boxplots are at first perceived as pictures, and gradually, as statistical understanding deepens, the students start to decode the pictures, and finally they begin to make judgements on and to argue about relevant features of the data. The links among fluency of decoding boxplots, ability to evaluate and form judgements, and instruction method are unknown. Understanding statistical language may also be a factor for some students who do not have English as their first language.

Fourth, when considering the connection between the students' and teacher's views, the method of instruction may be affecting students' level of performance and the views they adopt. The teacher used the traditional method of transferring back-to-back stem-and-leaf plot information onto boxplot representations. From that moment the students reasoned only with boxplot representations. Research suggests that students should be scaffolded to reason with boxplots through keeping the data, in dotplot form, under the boxplots (Bakker, Biehler, and Konold, 2005). The abrupt transition from stem-and-leaf plot to boxplot in instruction may be reflected in the referent view of half the students, which is largely focussed on naming the groups, and the fact that over half the students appear to be shape comparison describers or points readers. In other words, they reason as though there are no underlying data. Also, the teacher's referents are similar to the students. Furthermore, the game of informal inference is not made explicit to the students resulting in one teaching episode playing *game one* and *two*. The sampling view is not present in the students' responses, which may not be surprising since sampling reasoning is only communicated verbally by the teacher to the students. This verbal communication was not recorded by the students in their books and questions need to be raised about the influence of the written word on students' learning. The most important aspect, however, was that the students were not given opportunities to have experiences involving sampling variability and sample size effects. To develop students' inferential reasoning from distributions instruction needs to address and build concepts about sampling behaviour (Pfannkuch, 2005a). Finally, the signal view is present in one third of the students' responses. Since the teacher communicates this view verbally, visually, and in the written form this suggests that the instruction was effective in drawing students' attention to the middle 50% of data. Konold and Pollatsek (2002) note that students intuitively summarise data around a middle interval and therefore the teacher may have been tapping into and building onto this intuition. The question arises about how instruction can develop this view towards students viewing the median as a signal or being representative of the data set, and developing concepts about confidence intervals for the median.

Students' cognitive development and the method of instruction are intertwined. Research, however, needs to focus on how to develop students' inferential reasoning in a way that will lead them to formal inference with all types of graph comparison including co-variation. This small study suggests that improvement in inferential reasoning may depend upon more awareness of the multiple views taken when reasoning with boxplots, developing teacher and student talk, keeping data under the boxplots for as long as possible, and giving more opportunities to students to experience sampling behaviour.

REFERENCES

Bakker, A. (2004). *Design Research in Statistics Education: On Symbolizing and Computer Tools.* Utrecht, The Netherlands: CD-ß Press, Center for Science and Mathematics Education.

Bakker, A., Biehler, R., and Konold, C. (2005). Should young students learn about boxplots? In G. Burrill (Ed.), *Curricular Development in Statistics Education. International Association for Statistical Education (IASE) Roundtable*, Lund, Sweden, 28 June-3 July, (pp. 163-173), http://www.stat.auckland.ac.nz/~iase/publications.

Biehler, R. (2004). Variation, co-variation, and statistical group comparison: Some results from epistemological and empirical research on technology supported statistics education. Paper presented at the 10th International Congress on Mathematics Education, July, 2004, Copenhagen.

Friel, S., Curcio, F., and Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-159.

Key Curriculum Press Technologies. (2000). *Fathom*. Emeryville, CA: Author.

Konold, C. and Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259-289.

Pfannkuch, M. (2005a). Probability and statistical inference: How can teachers enable learners to make the connection? In G. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 267-294). New York: Springer.

Pfannkuch, M. (2005b). Informal inferential reasoning: A case study. In K. Makar (Ed.), *Reasoning About Distribution: A Collection of Research Studies. The Fourth International Forum on Statistical Reasoning, Thinking, and Literacy*, 2-7 July, 2005, Auckland, New Zealand [CD-ROM].