



International Electronic Journal of Mathematics Education

Volume 2, Number 3, October 2007

www.iejme.com

YEAR 11 STUDENTS' INFORMAL INFERENTIAL REASONING: A CASE STUDY ABOUT THE INTERPRETATION OF BOX PLOTS

Maxine Pfannkuch

ABSTRACT. Year 11 (15-year-old) students are not exposed to formal statistical inferential methods. When drawing conclusions from data, their reasoning must be based mainly on looking at graph representations. Therefore, a challenge for research is to understand the nature and type of informal inferential reasoning used by students. In this paper two studies are reported. The first study reports on the development of a model for a teacher's reasoning when drawing informal inferences from the comparison of box plots. Using this model, the second study investigates the type of reasoning her students displayed in response to an assessment task. The resultant analysis produced a conjectured hierarchical model for students' reasoning. The implications of the findings for instruction are discussed.

KEYWORDS. Statistics Education Research, Box Plots, Secondary Students, Informal Inferential Reasoning.

INTRODUCTION

Traditionally, up to Year 11 (15-year-old), New Zealand 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 (17-year-old), 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 instructional 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 becomes imperative rather than something to be avoided. A challenge for research is to understand the nature and type of reasoning used by students when making informal inferences from sample distributions about population distributions.

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. There is a need to understand students' informal inferential reasoning about many different tasks, but this study

focuses on the comparison of box plot distributions. In order to achieve the aim of attaining a greater understanding of students' reasoning when comparing box plots, a first study analyzed a teacher's reasoning, from which a model was developed. The second study, which is the main focus of this paper, then analyzed her students' reasoning on an assessment task using her model of reasoning.

With continuing changes in technology the statistics discipline has been inventing new ways of visualizing and exploring data. For example, Tukey (1977) invented box plots as a way of visually comparing the centers and spreads of batches of data. Such graphical techniques have filtered down to the education system and hence for the last twenty years box plots have been introduced to Year 10 (14 year-old) students in New Zealand. Teachers have always taught box plots from a descriptive perspective but recent changes to the Year 11 national assessment assume that students will draw conclusions from the comparison of box-plot distributions using visuo-analytical reasoning. The question arises as to what type of reasoning should be expected from these students.

REVIEW OF RELEVANT LITERATURE

There is limited research on how students reason from and interpret box plots. Bakker (2004) noted that the interpretation of box plots is conceptually demanding, since information is obscured, condensed, and summarized and statistical notions such as medians and quartiles are incorporated into the graph. Both Friel (1998) and Biehler (2004) reported that students tended to reason with and compare the five-number summary cut-off points when dealing with box plots. They theorized that the box plot's visual representation seems to lead students to focus intuitively on comparing cut-off points. In an earlier study Biehler (1997) conjectured that thinking about statistical summaries only with regard to their value rather than being properties of distributions may be a barrier to students when developing data analysis skills. Furthermore, he noted that "a conceptual interpretation of the box plot requires at least an intuitive conception of varying "density" of a data" (p. 178). In his 2004 study he confirmed that students, when reasoning with box plots, did not tend to comment about spread and the notion that the median was representative of the data set was difficult for students to understand. Biehler also found that students did not exhibit what he termed a "shift view" where the majority of the data appears to shift positions from one data set to the other, nor did they have intuitions about sampling variability, two elements of reasoning he considered essential for interpreting box plots.

Another problem is that statistics education has, until now, shied away from informal inference and there is no shared understanding of how to talk to students about graphs. Whether the research is focused on students' cognition by using innovative technology such as Fathom (Key Curriculum Press Technologies, 2000) or using students' own hand-drawn products, the problem of communicating and articulating the meaning of the statistical representations in classrooms remains difficult. Friel, Curcio, and Bright (2001) considered that research was

needed on understanding what it was about the nature of reasoning that made comparing data sets such a challenging task. They believed that graph comprehension involved an interplay between visual shapes, visual decoding, judgment, and context. Although Friel, Curcio, and Bright (2001) identified being able to speak about graphs as one of the behaviors that indicated students had developed a graph sense, Bright and Friel (1998) suggested that research was needed on how students and teachers talked and thought about graph representations. Biehler (1997) suggested that teachers and students are unsure about how to talk about graphs, since he found that when they do talk they use imprecise language. Furthermore, informal inference involves reasoning with data and justifying the conclusion in a process that is similar to presenting an argument (Bakker, Derry, & Konold, 2006).

The following section briefly describes the first study (see Pfannkuch, 2006 for a fuller account) that attempted to understand how one teacher talked about and reasoned with box plots.

STUDY ONE

The research described in this paper is part of a larger five-year project that is concerned with developing Year 11 (15 year-olds) students' statistical thinking based on the Wild and Pfannkuch (1999) framework. In the first year of the project, informal inferential reasoning was identified as a problematic area in the research classroom. Focusing on the comparison of box plots, the video-tape data of the classroom teaching revealed that the teacher in only one instance out of a possible eight opportunities communicated and wrote down how she would draw a conclusion from such plots (Pfannkuch & Horring, 2005). In an open-ended questionnaire, over half the students identified that they did not know how to draw evidence-based conclusions. An analysis of student responses to an assessment task requiring the drawing and justifying of inferences from the comparison of box plots revealed the following strategies:

1. Ninety percent of the students compared corresponding five-number summary statistics (e.g., lower quartiles of both groups) and 50% of them compared non-corresponding five-number summary statistics (e.g., lower quartile of one group with upper quartile of the other group), which is interpreted to be a "summary" element of reasoning;
2. Fifty percent of the students mentioned the difference in the ranges, a basic "spread" element; and
3. Thirty percent of them showed a very basic "shift" element of reasoning (Pfannkuch, 2005).

Realizing that drawing conclusions from the comparison of box plot distributions was not an easy task, the researcher and five statisticians met to discuss the type of reasoning that could be expected for informal inference for these students. Since Year 11 students had not been exposed to ideas of sample versus population or of sampling variability and the effect of sample size, the group conjectured that perhaps students should work with clear-cut comparisons that

had similar spread, no unusual patterns, and samples sizes of 30 elements. Even though comparing samples of size 30 could lead to incorrect conclusions being drawn using visuo-analytic reasoning (Pfannkuch (2005), there was the competing consideration that these box plots were being drawn by hand. Hence an insolvable conundrum existed when specifying what learning experiences were appropriate. The statisticians and researcher also suggested dot plots should be kept with box plots in order to make the link between representation and data more concrete (Bakker & Gravemeijer, 2004; Carr & Begg, 1994) and gave ideas on how students could experience sampling variation (Pfannkuch, 2005). After consideration of this information, the teacher decided that she wanted to deal with the inherent messiness of data where clear-cut decisions were not obvious and that her goal for the implementation of the unit in the following year would be communicating to students her reasoning processes when making informal inferences from the comparison of box plots.

From three teaching episodes on the comparison of box plot distributions, a qualitative analysis of the teacher's communication extracted ten elements of reasoning. These elements of reasoning adopted by the teacher are briefly summarized in Figure 1 (for a fuller account, see Pfannkuch, 2006).

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

Figure 1. Abstracted model of teacher's reasoning from the comparison of box plots

The eight elements of reasoning for comparing box plots are non-hierarchical, are interdependent but distinguishable, and are moderated by two other elements. This means that

the moderating elements, the referent and evaluative elements, are contained within each of the eight reasoning elements. For example, for the shift element, “the female graph is slightly higher than the male graph” the referents are “female” and “male” and the evaluative element is expressed by the use of the word “slightly”, as the strength of the evidence is assessed.

The goal of the teacher was to make an inference about populations from samples through comparing distributions and to justify that inference. Since informal inferences were being drawn from the comparison of two box plot distributions, the teacher used visuo-analytic thinking. She gradually built up, in her communication, the multifaceted ways in which she looked at and interpreted the comparison of the data sets. For the purposes of making comparisons with the students’ reasoning, in the second Study these elements will be briefly elaborated upon apart from the individual case reasoning element, which is not relevant to this age group, since the students did not identify outliers when comparing box plots.

To briefly illustrate the hypothesis generation, summary, shift, spread and explanatory elements consider the teacher’s written conclusion for a teaching episode where male and female pay was being compared using box plots (Figure 2(a)). The teacher constructed the box plots with the class and then discussed them. This written conclusion is not indicative of the teacher’s reasoning used in class but is employed here as an illustration of the elements of reasoning. The statements are numbered 1 to 4 for reference in the discussion. Note that LQ is an abbreviation for lower quartile and UQ is an abbreviation for upper quartile.

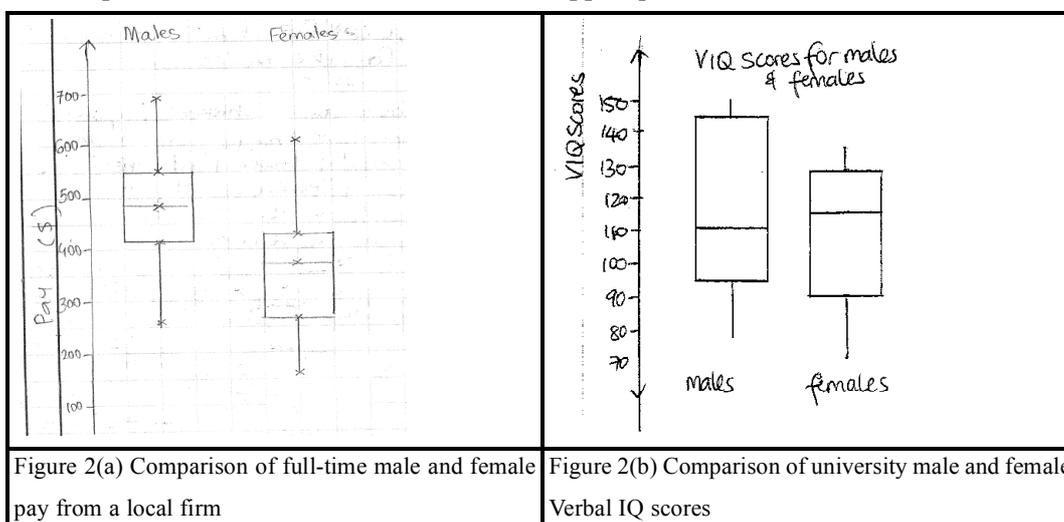


Figure 2. Comparison of box plots discussed in class

1. Each of the statistics (median, UQ, LQ) for females is lower than for the males.
2. The female graph is clustered between the median and the UQ.
3. The box and whisker graphs overlap but the female graph is generally lower than the males’ graph.
4. Overall, it appears that the females earn less than the males.

In the first statement, the teacher compared the median, upper quartile, and lower quartile of female pay with those of the male pay and hence she compared some corresponding five-number summary points. In her conversation she stated, “25% of females earn more than 25% of males” and therefore compared and interpreted non-corresponding summary points, that is, the female pay upper quartile with the male pay lower quartile.

In the second statement she referred to one aspect of the spread of the data for female pay, that is the distance between the median and upper quartile is short compared to the distance between the other quartile divisions, but she did not compare the spreads between male and female pay. In her spread element of reasoning the distinction between comparing variability within and between box plots was vague.

The third statement encompassed two reasoning elements, the signal (Konold & Pollatsek, 2002) and shift. In the signal element, when teaching, the teacher used the middle 50% of data as the rough signal amongst the noise. She compared 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 centers are compared relative to the variability. For the latter part of the statement, the teacher discussed with the class how she looked at the box plots as a whole to determine their relationship to each other and, hence, she used the shift element of reasoning.

The fourth statement is an example of hypothesis generation reasoning. She spent time discussing with students that “males earn more than females”, which was an inappropriate statement, since it was not true for every data value. In addition, she verbally expressed that “males tend to earn more than females, on average”. However, she did not record this language in her written conclusion. The explanatory element of reasoning was only expressed orally when she considered whether the findings made sense with what she and the students knew about female and male pay and whether there was a possible alternative explanation for the difference in pay, such as the position held being a factor rather than gender.

To illustrate the evaluative, referent, and sampling reasoning elements, consider the following abbreviated scenario from a teaching episode where the teacher compared university male and female students' Verbal IQ, which she referred to as IQ (Figure 2(b)). In this teaching episode the teacher gave the data and box plots to the class, since the focus of the lesson was on interpreting the plots.

She said: “I’ve got some conflicting information, the median for females show that they 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: “Based on these data values we are not certain that males have a higher IQ.” However, in response to a student who queried this

statement with, “but, couldn’t you say, from the graph, that males do have a little bit higher IQ than females?” She added: “there is some evidence to suggest that males have a higher IQ for these University students.”.

For this particular scenario the teacher spent a lot of time weighing the evidence before making a decision that she was unable to state whether males had a higher IQ, thereby demonstrating the evaluative element of reasoning. Unfortunately, in response to a student, she added that males had a higher IQ, indicating a conflict in her sampling reasoning. Her first written statement drew a conclusion about populations whereas her second statement drew a conclusion about the samples.

In relation to this, it should be mentioned that Pratt (2005) observed that these students needed to be aware of the game being played. He believed a large part of the students’ difficulty in understanding the sampling reasoning being used was that the teacher was not making the game being played explicit to the students. These students may have believed they were reasoning only about the data under consideration, which Pratt referred to as *game one*, whereas the teacher believed that the data were a sample from a population, which Pratt called *game two*. It is the playing of *game two* that will lead students towards formal inferential reasoning. Conflicting evidence led to a conflicting decision by the teacher, notably a game two decision about the populations, followed by a game one decision about samples. Even though, in this teaching episode, the teacher asked the students to imagine what the graphs might look like if another sample of people was chosen, or if the sample size of one plot was similar or much smaller than the other plot, her sampling reasoning may have eluded the students, since they had not been previously exposed to concepts such as samples, populations, and sampling behavior.

The referent element involves a constant back-and-forth switching between the visual symbol system, the box plot, and the concepts and ideas to which it refers. For example, other reference systems are the imagined dot plots underneath the box plots, or the statistical measures such as the median or the data measures. For this scenario the teacher’s referents were: males, females, IQ, median, and data values. The analysis of the teacher’s use of referents indicated that her language did not seem to sufficiently convey the underlying plot. For example, in her written conclusion, as discussed above for Figure 2(a), her main referents were male and female.

Study 1 resulted in abstracting elements of reasoning (Figure 1) that the teacher used when interpreting box plots. These reasoning elements showed the multifaceted richness of the conversation she communicated to her students. From this analysis many questions arose about the impact of her reasoning on the students: Would her students display similar reasoning? Would they use the same reasoning elements? Such questions gave rise to the following Study 2, in which the responses of her students to a box plot assessment task are analyzed. Based on the assumption that the students would imitate or be enculturated into the teacher’s way of reasoning, a decision was made to interpret the students’ responses in terms of the abstracted model of the teacher’s reasoning developed in Study 1.

STUDY TWO

Study 2 builds on the results from Study 1. As mentioned above this study is part of the long-term project on developing teachers' and students' statistical thinking in one school. Because of resource and time constraints data are gathered on one teacher and her class. The following research question is addressed in Study 2: Using the teacher's model of reasoning (Figure 1), what reasoning do her students articulate in an assessment task when drawing informal inferences from the comparison of box plots?

Method for Study Two

The research method is developmental in that an action-research cycle is set up whereby the teacher, researcher, and students identify problematic areas. The action-research method is based on the ideas of Gravemeijer (1998), Wittmann (1998), and Skovsmose and Borba (2000). The development of statistical thinking is about making changes and transformations in the classroom based on the notion that there should be an evolutionary development of a living system (Wittmann, 1998). The teacher and researcher visualize how the current situation might be changed, identify problematic situations based on research data, theorize and anticipate possible student learning trajectories, and explore alternatives to create an imagined situation (Skovsmose & Borba, 2000). The imagined situation eventually resides in a teaching unit. To move from the current situation the teacher must implement the teaching unit, the arranged situation, upon which, through a research analysis of the teacher's dialogue with students, student assessment tasks and responses to an open-ended questionnaire, critical reflections are produced on the teaching and learning process.

It is a collaborative research development, which may be viewed as a learning system in which all participants are learners and where there is a constant dialogue amongst the participants (Begg, 2002). The development is similar to a theory-guided bricolage as described by Gravemeijer (1998) except that the teachers develop the instructional activities rather than the researcher. Therefore the cycle of research is:

- First, to understand and evaluate the development of statistical thinking in the current situation with respect to teaching, learning, and assessment.
- Second, to collaborate with teachers in designing and testing a teaching unit intended to enhance students' statistical thinking.
- Third, to research the process of implementation with respect to teaching, learning, and assessment.

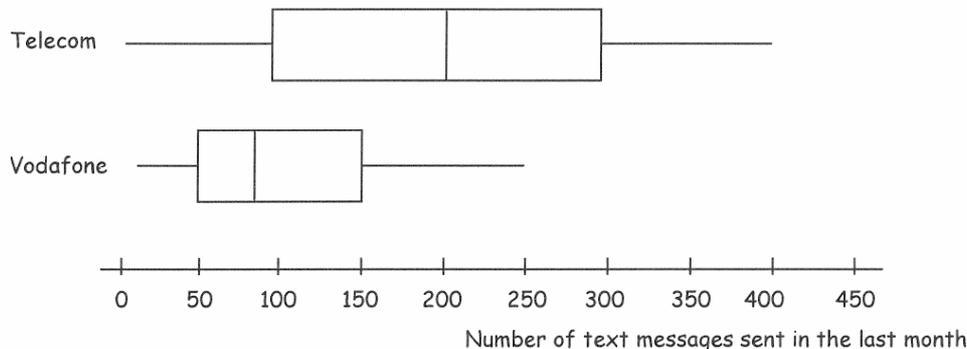
Since the research is conducted within a real course, there are constraints on its implementation, such as teaching time, availability of resources, and the fact that the students are working towards a national qualification.

The school in which the project is based is a multicultural, girls' secondary school. In the study class of 29 students, 40% were 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 box plots. 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 box plots. No technology was available to the students or teacher.

The teacher is 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 task.

There are two main Cell phone companies. They both 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 (approximately equal numbers of Vodafone and Telecom users). She has displayed the results using Box and Whisker Plots.



1. What is the Interquartile Range for the number of text messages sent in the last month for each of the phone companies?

Telecom _____ Vodafone _____

2. a) Loren asked the question: "Which phone company users generally sent more text messages in the last month?" Respond to her question.

b) How does the graph show this? Make 3 statements to explain.

Figure 3. Student assessment task

This Study focuses on the student assessment responses to a task involving comparison of box plots and compares them to the teacher's communication as described in Study 1. The researcher analyzed the student assessment responses but no independent researcher was available to check the reliability of the data classifications. The teacher, however, did confirm the overall interpretation of the student assessment data.

The Assessment Task: The student assessment task is illustrated in Figure 3. This task was similar to other class activities, and was part of a larger assessment on the statistics unit, which was sat by students for a national qualification.

Results for Study Two

For Question 1 of the assessment task (Figure 3) all students were able to identify and read the upper and lower quartiles from the Telecom and Vodafone box plots and calculate the interquartile ranges of 200 and 100 respectively.

Question 2(a) invited the students to identify the phone company from the data presented. Question 2(b) became the focus of the analysis since students must justify their reasoning for their response to the question. By using a spreadsheet the student responses were first sorted into each reasoning element. Once they were sorted, a three-level hierarchy of performance emerged from a qualitative analysis of each reasoning element, from which qualitative descriptors were developed (Figure 4). These level descriptors will be illustrated in the next three pages. From the descriptors for each of the three levels, an overall classification was given that seemed to describe the type of reasoning that was prevalent within each level, that is shape comparison describer, decoder, and assessor. Since some students did not respond to Question 2(b), gave responses that were inappropriate, or were minimal in terms of reasoning, a pre-cursor level was added (point decoder) to describe those students who could calculate the interquartile ranges in Question 1.

Using the qualitative descriptors (Figure 4), each student response for each reasoning element was assigned a level. After the students were coded on one of three levels for each element of reasoning, the following criterion was developed to nominate their overall level of reasoning: To score an overall level of X , student must have more than one reasoning element present (i.e., summary, spread, shift, signal) and obtain at least level X for three elements (includes the moderating elements), otherwise overall score was $X-1$.

Elements of reasoning	Point Decoder Level 0	Shape Comparison Describer Level 1	Shape Comparison Decoder Level 2	Shape Comparison Assessor Level 3	Total students using element
Summary	Identifies the 5-number summary points	Compares two or more corresponding 5-number summary points including the median.	Compares medians only. Compares non-corresponding 5-number summary points but does not interpret.	Compares non-corresponding 5-number summary points and interprets.	26
Spread		Compares spreads of visual shapes (lengths, spaces in boxes) but does not decode.	Compares and talks about spread, refers to range. Compares local regions/densities.	Compares and refers to the type of spread/densities locally and globally (e.g. even, clustered).	17
Shift		Compares and refers to the shift qualitatively for the whole shape.	Compares and refers to the type of shift (e.g., non-uniform).	Compares and refers to the shift of the majority.	4
Signal		Compares the middle groups' visual shapes (boxes) in relation to each other but does not decode.	Compares the middle groups and decodes by referring to the data.	Compares the overlap of the data of middle groups.	10
Moderating elements					
Referent		Partial context referent – refers to names of groups only. Statistical measure referent (e.g., median)	Context/partial data referent – refers to name of the data (e.g. text messages) and/or talks about the data only in one reasoning element.	Data referent – refers to and talks about the data or data plot underneath the box plot in two or more reasoning elements.	
Evaluative		Compares appropriate difference by description (e.g., higher, further up).	Ascertain strength of the evidence for appropriate comparisons (e.g., a lot higher, much further along).	Ascertain strength of the evidence and then weighs evidence (e.g., even though they overlap, not too much overlap).	
Total students attaining overall level	8	10	9	2	

Figure 4. Level descriptors for student reasoning from comparison of box plots including summary of student overall attainment level

From Figure 4 it should be noted that some reasoning elements were not present in the students' communication. The absence of the hypothesis generation element can be explained by the fact that Question 2(a) gives the hypothesis and hence students would not be expected to write this down. The absence of the sampling and explanatory elements of reasoning could be

explained by either the fact that the teacher did not write these reasons down on the whiteboard, even though she spent a lot of time in class using them in her own reasoning, or by the fact that conceptual foundations had not been built for students to understand these reasoning elements.

Some student responses to Question 2(b) of the assessment task are now discussed to illustrate the reasoning levels. Student 11 (Table 1) is an example of a point decoder. A point decoder is able to read and compare corresponding five-number summary points, and calculate learnt formulaic procedures, such as the interquartile range, which all 29 students accomplished in Question 1. Student 11 compares the maximum values and incorrectly attempts to compare the median number of text messages for Telecom with the maximum value for Vodafone. Her reasoning is entirely situated within comparing the five-number summary cut-off points and there is no sense that she is comparing shapes or distributional features.

Table 1. Example of a point decoder

Student 11 Quote	Designated Level
The highest value of the Telecom phone company is 400 and Vodafone is only 250. The median of Telecom phone company is overlapping the highest value of Vodafone company [Note: incorrect statement].	Level 0 Summary
Moderating elements within quote	
Referents: Telecom, Vodafone, statistical measure	Level 1 Referent
Evidence: inappropriate to weigh evidence by comparing maximum values	Level 0 Evaluative

Shape describers seemed to be reasoning only with the “pictures”, entirely ignoring statistical concepts and the plots of the data represented by the box plots. Student 3 is an example of a shape comparison describer (Table 2). Ignoring Student 3’s summary element of reasoning, note that, in her reasoning, she is describing the difference between two similar box images and that her statistical language associated with comparing box plots is non-existent, though she does use the term “overlaps”.

Table 2. Example of a shape comparison describer

Student 3 Quote	Designated Level
Telecom has higher Median, Upper Quartile, Lower Quartile, lower Quartile and Max point.	Level 1 Summary
Telecom is more to the higher points of the scale.	Level 1 Shift
Telecom box overlaps Vodafone	Level 1 Signal
And the spaces in the Telecom box are bigger than that of Vodafone.	Level 1 Spread
Moderating elements within quote	
Referents: Telecom, Vodafone, statistical measures.	Level 1 Referent
Evidence: described only by “higher, more to the higher, overlaps.”	Level 1 Evaluative

Only two students were coded as being at Level 1 for every element of their reasoning. The other students displayed a continuum of levels such as Student 29 whose overall level was deemed to be a shape comparison decoder (Table 3). Shape decoders were 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. Student 29 was starting to use statistical terminology by comparing percentages of messages sent and she showed an awareness that the box plots were representing data.

Table 3. Example of a shape comparison decoder

Student 29 Quote	Designated Level
All the values for Telecom are higher (median, LQ, UQ).	
Over 75% of Telecom users sent more messages than 50% of Vodafone users.	Level 3 Summary
The boxes aren't overlapping that much.	Level 1 Signal
Moderating elements within quote	
Referents: Telecom users, Vodafone users, statistical measures, messages	Level 2 Referent
Evidence: weighed by indicating middle groups of data do not overlap that much.	Level 3 Evaluative

Shape assessors were using statistical language more fluently to describe the data, comparing appropriate features as evidence, and beginning to make judgments on relevant evidence. No student showed a high level of accomplishment in this level. Student 21 is an example of reasoning that is beginning to show signs of a shape comparison assessor (Table 4). In her response there is a strong sense that she is reasoning about data although she does not contextualize the data as text messages; statistical language such as “evenly spread out” and “clusters” appears; and in the signal element she compares the central group of data, weighs the evidence, and attempts to justify her reasoning.

Table 4. Example of a shape comparison assessor

Student 21 Quote	Designated Level
Each of telecom's data values (median, UQ, LQ) are higher than the corresponding Vodafone data values.	Level 1 Summary
Telecom's data is more evenly spread out, whereas Vodafone's data clusters between the median and lower quartile.	Level 3 Spread
The central bulk of the data for the two graphs does overlap, but Vodafone's upper quartile does not exceed above Telecom's median.	Level 3 Signal
Moderating elements within quote	
Referents: Telecom, Vodafone, statistical measures, data underneath box plots in more than two elements	Level 3 Referent
Evidence: weighed by indicating middle groups of data do overlap but ...	Level 3 Evaluative

Twenty-six of the students presented a summary element of reasoning whereby, at the very least, they compared some corresponding five-number summaries or cut-off points, but only three were coded at Level 3. Seventeen students used a spread element of reasoning, with twelve of them starting to use language associated with the spread of the data. Only four students illustrated a shift element of reasoning with two of them being classified at Level 3. Interestingly, ten students were categorized as having a signal element in their reasoning, while seven students used the term “overlap”, an element and term not expressed by the previous year’s students in a similar study (Pfannkuch, 2005). For over half the students, their only referents were the group label and statistical measures. For the evaluative element, eight students were assessing the strength of the evidence with statements such as: ‘Vodafone’s median is much lower than telecom’s median’ or “the median for Telecom was 200 whereas Vodafone was only about 70.” Four students weighed the evidence appropriately. The students’ elements of reasoning are a subset of the teacher’s oral reasoning but are the same reasoning elements she presented in the written form.

DISCUSSION

The research question addressed in Study 2 was concerned about the type of reasoning Year 11 students use when drawing informal inferences from the comparison of box plots. The question will be discussed in terms of the students’ reasoning, the conjectured level descriptors for student performance, and the connections between the students’ and teacher’s communication.

Since twenty-six students attempted a summary element of reasoning, Friel (1998) and Biehler’s (2004) contention that students intuitively argue with the five-number summary cut-off points is supported by this research. Biehler (2004) noted that, despite instruction, his students lacked a “shift view” and intuitions about sampling variability, which also seems to be the case with students in Study 2. But over half of these students did pay attention to the spread of the data, a facet lacking in Biehler’s students.

The conjectured overall descriptors of point decoder, shape describer, decoder, and assessor support Friel, Curcio, and Bright (2001) belief that comprehension of graphs involves visual shapes, visual decoding, and judgment. Since only eleven students were reasoning beyond a shape describer, Study two indicates that these students find it difficult to verbally express, describe, and justify conclusions from the comparison of box plots. The level of understanding demonstrated by these students confirms Bright and Friel’s (1998) and Biehler’s (1997) conjectures that students are unsure about how to talk about box plots, despite some hours of instruction. A challenge for research is to explicitly describe and understand the conceptual building blocks and the argumentation processes for informal inference that will later lead students to a sound comprehension of statistical inferential reasoning (see Ben-Zvi, 2006; Hammerman & Rubin, 2006; Rubin, Hammerman, & Konold, 2006).

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 box plots are at first perceived as pictures. Gradually, as statistical understanding deepens, the students start to decode the pictures, and finally they begin to make judgments on and to argue about relevant features of the data. The links among fluency of decoding box plots, ability to evaluate and form judgments, 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.

When considering the connections between the students' and teacher's reasoning, the method of instruction may be affecting students' level of performance and the reasoning elements they adopt. The teacher (Study 1) used the traditional back-to-back stem-and-leaf plot as a computational aid to transfer statistical summary information onto box plot representations and the data in the assessment tasks was only given in this format. From that moment her students (Study 2) could reason only with box plot representations. Previous research suggests that students should be scaffolded to reason with box plots through keeping the data, in dot plot form, under the box plots (Bakker, Biehler, & Konold, 2005), but dot plots were not provided in the assessment task. The abrupt transition from stem-and-leaf plot to box plot in instruction may be reflected in the referent element of reasoning, since half the students largely focused on naming the groups, and about two-thirds of the students appeared to be shape comparison describers or point decoders. In other words, they reasoned as though there were no underlying data. Also, the teacher's referents were similar to the students.

Furthermore, the sampling and explanatory elements were not present in the students' responses, which may not be surprising since this reasoning was only communicated verbally by the teacher to the students. Since students did not record this verbal communication in their books, 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 such as sample distribution, population distribution, and sampling behavior (Pfannkuch, 2005; Saldanha & Thompson, 2007). The explanatory element, which is also highlighted by Friel, Curcio, and Bright (2001) in terms of understanding the contextual frame of the data, is necessary for interrogating and drawing inferences from data. However, how students reason with this element is unknown.

The signal element, however, was present in one third of the students' responses. Since the teacher communicated this reasoning 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) and Bakker, Derry, and Konold, (2006) note that students

intuitively summarize 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 element of reasoning, so that students view the median as being representative of the data set, and help to develop intuitive 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 reasoning elements associated with box plots, developing student talk and argumentation processes, keeping data under the box plots for as long as possible, and giving more opportunities to students to understand the concepts of sample and population and to experience sampling behavior.

LIMITATIONS

There are three main limitations to this research. First, the study has only captured students' reasoning from box plot representation of distributions from one teacher's class. Second, students were not interviewed to determine whether other elements of reasoning were orally present. Third, one researcher categorized the elements and hence there is no triangulation from independent sources although the teacher did take the opportunity to assess the interpretation but did not make any changes. Hence, this research can only offer some insight into possible ways students may be reasoning informally and into possible pitfalls in the reasoning process. Therefore these findings remain speculative since the study is not representative and has a small sample size.

IMPLICATIONS FOR FUTURE RESEARCH

The research described in this paper is a collaborative process between the researcher and teachers. Consequently, the implications for teaching from this research resulted in the teachers agreeing to change the way they traditionally taught. At the Year 10 level the research is now focusing on presenting dot plots and box plots simultaneously, giving teachers and students experience of sampling behavior with categorical and numerical data and with different sample sizes, distinguishing between sample and population distributions, and developing concepts of random sampling. Improving teachers' knowledge of statistics and their statistical reasoning and thinking are also important keys to improving students' reasoning and thinking. Therefore, future research needs to identify the types of content knowledge and pedagogical knowledge for teaching statistics that teachers need to help students be successful in developing informal ways of comparing distributions and in particular students' informal statistical argumentation.

REFERENCES

- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: CD-B Press, Center for Science and Mathematics Education.
- Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill, & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable* (pp. 163-173). Voorburg, the Netherlands: International Statistical Institute and International Association for Statistical Education. On line: <http://www.stat.auckland.ac.nz/~iase/publications>.
- Bakker, A., Derry, J., & Konold, C. (2006). Using technology to support diagrammatic reasoning about center and variation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.
- Begg, A. (2002). Rethinking curriculum: A developer's perspective. In B. Barton, K.C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific. Proceedings of the 25th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 123-130). Sydney, NSW: MERGA.
- Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.
- Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education. On line: <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 *Tenth International Congress on Mathematics Education*, Copenhagen, July, 2004.
- Bright, G., & Friel, S. (1998). Graphical representations: Helping students interpret data. In S. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K-12*, (pp. 63-88). Mahwah, NJ: Erlbaum.
- Carr, J., & Begg, A. (1994). Introducing box and whisker plots. In J. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics (ICOTS 4) Marrakech, Morocco*, July, 1994: Minneapolis, MN: University of Minnesota.

- Friel, S. (1998). Comparing data sets: How do students interpret information displayed using box plots? In S. Berenson, K. Dawkins, M. Blanton, W. Coulombe, J. Kolb, K. Norwood, & L. Stiff (Eds.), *Proceedings of the Twentieth Annual Meeting, North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 365-370). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-159.
- Gravemeijer, K. (1998). Developmental research as a research method. In A. Sierpiska and J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (Book 2, pp. 277-295). Dordrecht, The Netherlands: Kluwer.
- Hammerman, J., & Rubin, A. (2006). Saying the same (or a different) thing: How shape affects ideas about distribution in a software exploration environment. In A. Rossman, & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.
- Key Curriculum Press Technologies (2000). *Fathom*. Emeryville, CA: Author.
- Konold, C., & 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. (2005). 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. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27-45. On line: <http://www.stat.auckland.ac.nz/~iase>.
- Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable* (pp. 204-218). Voorburg, the Netherlands: International Statistical Institute and International Association for Statistical Education. On line: <http://www.stat.auckland.ac.nz/~iase/publications>.
- Pratt, D. (2005). *Personal communication at The Fourth International Forum on Statistical Reasoning, Thinking, and Literacy on Reasoning about Distribution*, Auckland, New Zealand, July 2-7, 2005.
- Rubin, A., Hammerman, K., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.

- Saldanha, P., & Thompson, P. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, this issue.
- Skovsmose, O., & Borba, M. (2000). *Research methodology and critical mathematics education*. (Publication No. 17). Roskilde, Denmark: Centre for Research in Learning Mathematics, Roskilde University.
- Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223-265.
- Wittmann, E. (1998). Mathematics education as a 'design science'. In A. Sierpiska and J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity* (Book 1, pp. 87-103). Dordrecht, The Netherlands: Kluwer.

Author : **Maxine Pfannkuch**
E-mail : m.pfannkuch@auckland.ac.nz
Address : Department of Statistics, The University of Auckland
Private Bag 92019, Auckland, New Zealand
Phone : 64-9-3737599
Fax : 64-9-3737018