Paper Presented at the AARE Annual Conference, Brisbane, 2002

School Students' Understanding of Stacked Dot (or Line) Plots

Jane M. Watson and Ben A. Kelly

University of Tasmania

Abstract

This study considers students' understanding of a relatively new form of graph in the data-handling curriculum: the stacked dot plot, or as it is more commonly known, the line plot or dot plot. Students in grades 5 to 10 were presented with various forms of a task asking for comparison of two stacked dot plots, one with usual scaling of the base line and the other with gaps for zero-values omitted. Of interest were students' abilities to interpret the information as presented in the graphs, to put the information in context, and to distinguish the statistically appropriate form of graph with an explanation. Performances across grades and forms of the task are compared. For a subset of students in grades 5, 7, and 9 performance is compared for one form of the task before and after a unit of study on chance and data. Discussion includes implications for the classroom in relation to current interest in statistical literacy across the curriculum.

Introduction

Line plots, dot plots, or stacked dot plots as they have been referred to by Konold, Robinson, Khalil, Pollatsek, Well, Wing, and Mayr (2002), have had a presence in the data-handling school curriculum for less than 20 years. It was the *Quantitative Literacy Series* that first introduced them to secondary teachers as line plots in a graphing activity based on the number of medals won at the 1984 Winter Olympic Games (Landwehr & Watkins, 1986, pp. 1-2). Following this, Russell and Corwin (1989), in the *Used Number Series* used line plots to teach primary students a "quick way to show the shape of data" (p. 17) as part of an introduction to "data analysis - collecting and displaying". The terminology is further used in the *Elementary Quantitative Literacy Series* (Bereska, Bolster, Bolster, & Schaeffer, 1998, 1999) with the following definition provided for teachers: "A line plot is a graph that indicates the location of data points along a segment of the real number line" (p. 121; p. 149). The National Council of Teachers of Mathematics (NCTM) *Principles and Standards* (2000) also refers to line plots in the section on Data Analysis and Probability for grades 3 to 5. A recent text book for mathematics preservice secondary teachers (Dossey, McCrone, Giordano, & Weir, 2002) uses the term dot plot to refer to these same representations. The use of the term dot plot at the secondary level probably reflects a usage adopted by some statisticians (e.g., Chambers, Cleveland, Kleiner, & Tukey, 1983; Cleveland, 1993; Wilkinson, 1999). Chambers, et al. (1983), were some of the earliest to use such terminology but referred to the graph as a one-dimensional scatter plot with stacking. Wilkinson mentions stacking dots but does not go on to use the term in the title of this paper. In advanced circles, however, many adjectives are beginning to be used to describe dot plots, e.g., Cleveland uses multiway dot plots and Wilkinson uses asymmetric, symmetric, and histo-dot plots.

In their basic form, these plots are a quick and efficient means of displaying and summarising data. They are easy to construct, and can be created by most students. Stacked dot plots make interpreting and constructing graphs accessible to a wide range of students, enabling data analysis to become manageable from a young age. They give students the ability to familiarise themselves quickly with a data set, to identify the concentration of data such as in clumps, to acknowledge the absence of data, and to pinpoint unusual data and outliers (NCTM, 2000). They are not, however, a part of the traditional school statistics graphing curriculum and anecdotal evidence suggests that many high school teachers do not consider them legitimate, perhaps due to the confusion with line graphs (cf. Figure 1). This dilemma prompted Konold to advocate the new nomenclature (personal communication, 26 March, 2002), choosing the extra adjective based on usage by Wilkinson (1999), and to begin to use it (Konold et al., 2002).

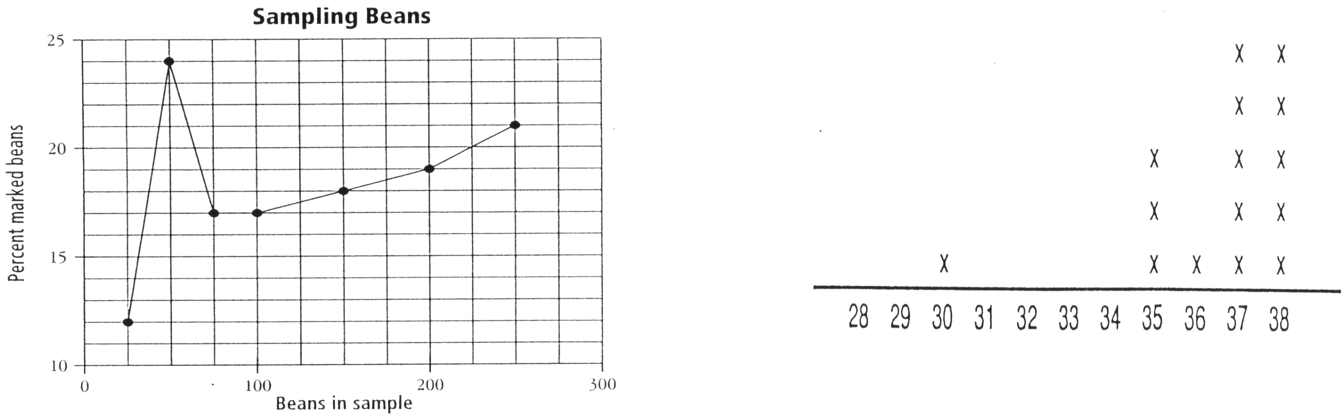


Figure 1. Difference between a line graph (Lappan, Fey, Fitzgerald, Friel, & Phillips, 1997, p. 64j) and line Corwin, 1989, p. 17).

Little research has been carried out, however, into students' reactions to and understanding of stacked dot plots. A case study in Russell, Schifter, and Bastable (2002) and an analysis by Konold and Higgins (2002) suggested that when students were asked to draw stacked dot plots from raw data, the issue of what to do about values with no observations (0-value data) was a significant problem. Although the definition suggests that the base line of the graph is the real number line with no gaps, children with little experience, and who are initially given whole number data, may not appreciate the need for including all values as part of a line segment. Further research by Konold et al. (2002) into students' descriptions of data sets represented by stacked dot plots found a preference for discussing typicality in terms of "modal clumps" of stacked dots rather than in terms of means or medians. A "modal clump" was defined as a range of data in the heart of a distribution of values.

This research, conducted as part of three studies, explored students' reactions to stacked dot plots in a test setting using stacked dot plots that were actually created by students (Konold & Higgins, 2002). Five questions were of interest. (1) What are the levels of understanding displayed for the stacked dot plot task? (2) Are there differences in performances for different forms of presentation of the task? (3) Are there differences in performance at different grade levels? (4) Does instruction in chance and data improve performance on the stacked dot plot task? (5) Following the observations of Konold et al. (2002), is there evidence of students providing unsolicited comments relating to modal clumps in a stacked dot plot?

Sample

Students in Study 1 were from 10 government schools in the Australian state of Tasmania in grades 5, 7, and 9 who were participating in a larger study investigating school students' understanding of statistical variation. They were given a survey including a task based on stacked dot plots. Study 2 consisted of students from 5 Tasmanian Catholic schools who were participating in a trial of a Statistical Literacy survey. These students were further divided into two groups, each answering a slightly different version of the stacked dot plot task. The numbers of students in each grade, completing each version of the survey, are shown in Table 1. It is not known whether students in Study 1 and Study 2 had seen stacked dot plots before but they were not part of the suggested curriculum in chance and data.

Table 1

Number of Students per Grade and Study

		G5	G6	G7	G8	G9	G10	TOTAL
Study 1	Form I	156		162		172		490
Study 2	Form II	85	29	61	44	40	52	311
	Form III	47	93	46	48	59	43	336
Study 3	Form I	72		79		76		227

Study 3 consisted of a sub-sample of the students in Study 1 in five of the 10 schools who were re-tested with the same survey as Study 1, after completing specialised lessons focusing on chance, data handling, and statistical variation. The students in grade 5 were certainly exposed to stacked dot plots as part of their unit of work and it was recommended to teachers of grade 7 and 9 that stacked dot plots be included in the lessons provided. Based on the journals kept by the teachers in grade 7 and 9, five of nine classes were exposed to stacked dot plots in their lessons. Although four teachers mentioned graphing, they did not specify that stacked dot plots were used for this purpose. Table 1 contains the numbers of students in Study 3 in grades 5, 7, and 9.

Survey Items

For Study 1 and 3, the task shown in Figure 2, adapted from Konold and Higgins (2002), was given to students in grades 5, 7, and 9. The first two parts to the item asked students to interpret two stacked dot plots showing how many years students from a class had lived in their town. The third question asked students to evaluate which of the two stacked dot plots told the story better. Although including identical data values in both plots, the scale of the first plot included only data values actually occurring in the data set, whereas the second plot used all possible data values in the range.

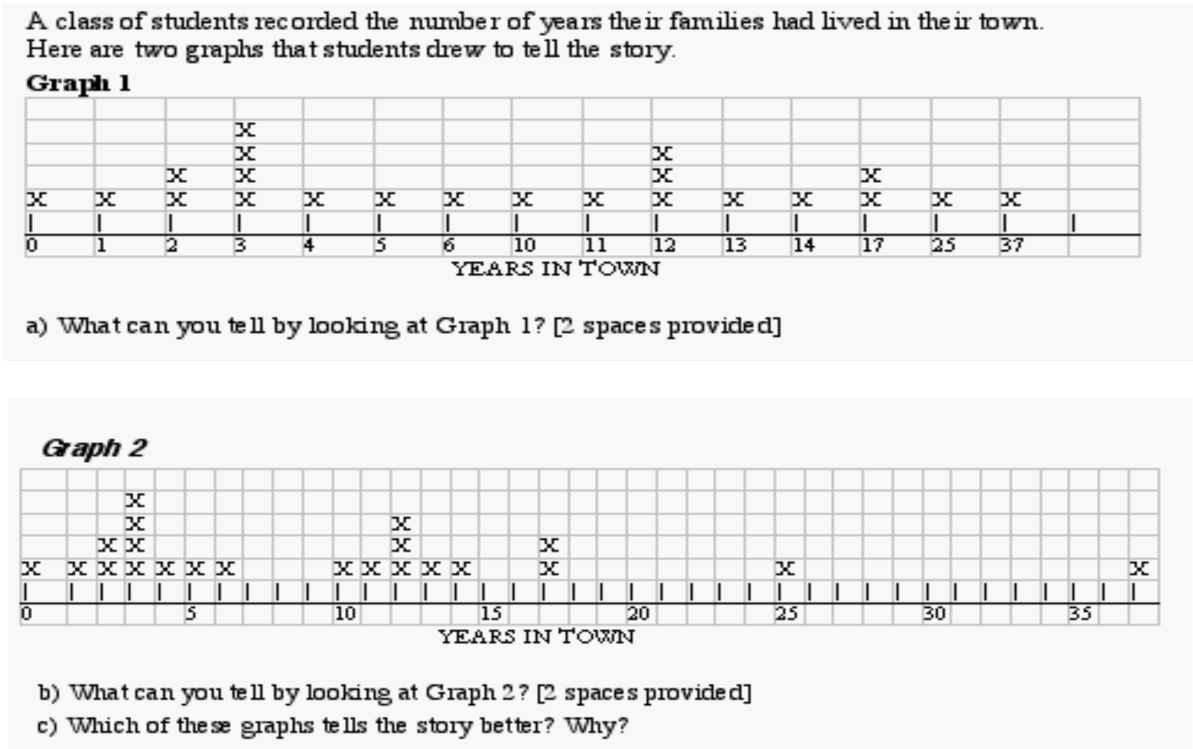
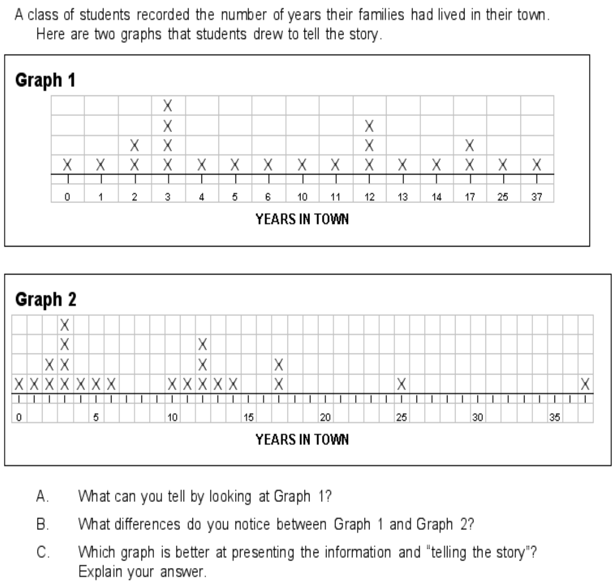


Figure 2. Form I of the stacked dot plot task used in Study 1 and Study 3.

For Study 2, the task from Figure 2 was slightly modified (see Figure 3), and for Forms II and III, a different stacked dot plot was presented first as Graph 1 and questions initially directed to it. On Form II, for example, students were first asked what they could tell by looking at Graph 1 (the stacked dot plot that included only data points that actually occurred). They were then asked what differences they noticed between Graph 1 and Graph 2 (in this case the stacked dot plot that used all possible data points), and finally, which graph was better at presenting the information and telling the story and why. On Form III, the same questions were asked as for Form II, however, the students were presented with the stacked dot plot that used all the possible data points first as Graph 1 and the stacked dot plot that used only data points actually occurring as Graph 2 (see Figure 3). The reasons that this design was employed were to avoid repetitive responses or non-responses observed to Part b) in Form 1 and to explore the possibility that focussing and commenting on the "statistically appropriate" plot first, might produce different performance from students.



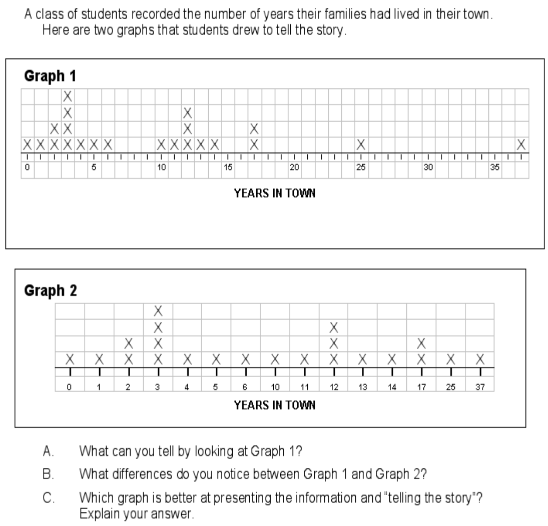


Figure 3. Form II (left) and Form III of the stacked dot plot task used in Study 2.

Analysis

Two aspects of responses were used in developing a coding scheme for responses. First was the statistical appropriateness of responses in terms of explaining issues related to the physical representation of the stacked dot plots (cf. Bereska et al., 1998) and to the context or story addressed by the data set in the task. Second was the structural complexity of the responses given the request for two statements in Form I (cf. Biggs & Collis, 1982; Collis & Biggs, 1991). Comments relating to reading the data directly from the stacked dot plot were assigned lower values in the hierarchical coding scheme than comments that linked the information in the plot to the context of the task. The latter responses appeared to the authors to reflect structurally more complex thinking.

All responses were entered into spreadsheets. For Studies 1 and 3, after agreement of the authors on expected levels of response, coding was carried out by the second author. The coding procedure clustered similar responses into separate codes based on the type of reasoning shown (Miles & Huberman, 1994). Both authors checked the consistency of coding, with discrepancies in coding resolved by discussion. For Study 2, the codings used in Study 1 were part of a code booklet supplied to a research assistant who scored the Statistical Literacy survey. Codes used for Part a) and b) of Form I were adapted for Part a) of Forms II and III and codes used for Part c) of Form I were adapted for Parts b) and c) of Forms II and III. Because only one reason was solicited for Part a) of Forms II and III, the criteria were less stringent, based on statistical appropriateness. For all forms of the task, all parts of the task were coded in four levels. The four coding levels for all parts of items reflected increasing understanding and sophisticated structure.

Means and standard deviations for each form of the question were calculated based on the coding and paired *t*-tests were performed on the pre and post conditions for Study 3.

Analysis of variance and *t*-tests were also used to assess differences in performance among grades and between Forms II and III of the tasks.

The evidence for student recognition of modal clumps was sought first in the responses to Form 1, Part b) and Form III, Part a) where the question asked what students could tell from the appropriately scaled stacked dot plot. Second it was sought from Part c) of all three forms in responses that provided reasons for selecting the appropriate plot. In each case percents of responses were recorded from the coding categories of appropriate responses.

Results

The first part of the Results will present examples of the levels of understanding for the three forms of the stacked dot plot task. Levels are equated with the codes assigned. This will include a summary of the overall frequency distribution of responses. Differences in overall performance among grades and for the three forms will then be noted, as will the difference after instruction for students answering Form I. Finally, evidence for student identification of modal clumps will be presented.

Development of Understanding

For Form I, a code of 0 was associated with responses with no discernable logical reasoning (e.g., "That it must be a nice town"). At Level 1, responses included data reading comments or if a summary statement was made, it was combined with an inappropriate statement. At Level 2, responses included one summary statement and one data reading statement, or a single summary statement with no other comment. At Level 3, responses provided two distinct appropriate summary statements. Examples are given for Parts a) and b) of Form I in Table 2.

Table 2

Examples for Part a) and Part b) in Form I of the Task with Percent in Each Level

Level		Examples for Part a) for Form I	Examples for Part b) for Form I
0	No response; idiosyncratic or inappropriate comment on interpretation or the type of graph	"Not sure."	"Not much."
		"3 is the biggest family."	"The most people were at 0.3."
		"It's easier to look at but it's not really good enough."	"The families were big but as time went on they went down."
		"It's set out differently to graph 2." **23.47%**	**36.33%**
1	At least one appropriate data reading comment, or	"One person has not lived on the town for 1 year. Most people have lived in the town for 1 year."	"Two people have lived here for 2 years. Eleven have lived here for 1 year."

	one appropriate summary comment accompanied by one inappropriate comment	"People who have lived in their house for 3 years is four people. People who have lived in their house 37 years is one." "12 years has 1 less than 3 years. There is only 1." **24.69%**	"Someone stayed there for 10 years. And some one stayed there for 1 year." "It tells you how many families lived in the town over 37 years. It tells you when people moved to a different town." **17.35%**
2	One appropriate summary, perhaps with an appropriate data reading comment	"A lot of people were here 3 years."	"Not many people lived here long."
		"That most people have stayed in town for different amounts of time. Four people have stayed for 3 years."**28.16%**	"4 families have lived there for 3 years. No families have lived there during 18, 19, 20, 21, 22, 23, 24." **28.98%**
3	Two appropriate summary comments	"That most people don't live in the town much after 12 years. The majority have live there for 10 years or less." **23.67%**	"Well, more families lived in their town around the 5/4 year mark. But 10 and 15 years is quite popular."**17.35%**

The percents of students responding at each level are also shown in Table 2 and highlight the differences across the four levels of responses for Parts a) and b). Part a) shows a lower percent of Level 0 responses, and a higher percent of Level 1 and Level 3 responses when compared to Part b). Overall, performance on Part a) in Form I was better than for Part b), with students giving more complex and statistically appropriate answers.

For Part c) of Form I, at Level 0 were responses with no logic, or a statistically inappropriate choice of graph (Graph 1 in Figure 2) with either no reasoning or idiosyncratic reasoning. At Level 1, responses included the statistically inappropriate choice of plot again, with reasoning that was personalised or based on the numbering on the scale or general graph lay-out. At Level 2, the statistically appropriate plot was chosen (Graph 2 in Figure 2), however, either no reasoning or non-statistical and / or personalised reasoning was given. Respondents who were indifferent to choosing the better plot, based on legitimate statistical arguments were also coded at Level 2. At Level 3, again the statistically appropriate plot was chosen but this time statistical reasoning and concise arguments were presented.

Examples of responses for each level are given in Table 3. Once again the overall percent of responses for each of the levels is included. As can be seen, there is a relatively stable trend across the four levels of responses, with slightly more responses occurring in Levels 1 and 2. Level 1 and 2 responses rely heavily on personal preferences to determine which graph is better (see Table 3), and state general comments about the data (sometimes incorrect) or graph lay-out. The lowest frequency response level was Level 3 and shows that fewer students responded with the statistically appropriate graph *and* appropriate reasons for their choice.

Table 3

Examples for Part c) in Form I of the Task with Percent in Each Level

Level		Examples
0	No response; idiosyncratic or incorrect interpretation of Graph 1	"Don't know - I can' t really tell, maybe 1."
		"Graph 1 - Because I do."
		"Graph 1 - Because the people lived there longer." **24.08%**
1	Support for Graph 1 with reasoning based on personal preference, the numbering on the axis, or the lay out	"Graph 1 - Because it is easier to look at and to count."
		"Graph 1 - Because you didn't have to count from 0-5-10-15 etc."
		"Graph 1 - Because the graph seems to be more spread out and realistic." **29.80%**
2	Support for Graph 2 or both with no reasoning or reasoning based on inappropriate interpretation or personal preference	"Graph 2 - Because I think that."
		"Graph 2 - Because it's got more people."
		"Graph 2 - Because with graph 2 you might live here say like 5 years and after that you might want to move then."
		"Graph 2 - Because it is easier to read and understand."
		"Both - Because they nearly tell you the same thing." **28.57%**
3	Support for Graph 2 with statistically appropriate reasons and implications	"Graph 2 - Because it shows you the years that no one lived in the town." "Graph 2 - Because it shows each year individually."
		"Graph 2 - Because graph 2 is more visual and you can notice the time difference in how long each family has been there more precisely." **17.55%**

For Part a) of Forms II and III, the criteria for Form I Parts a) and b), were simplified to reflect the requirement for a single response. The levels and examples responses for Part a) for each form are found in Table 4. The overall distribution of responses is also given. Examples of student responses were similar in content to those presented in Table 2. Form III reveals a higher response rate at Level 3 than for Form II. This is balanced by a higher response rate for Level 2 in Form II than for Form III. It seems that it was easier for students to summarise the data and respond optimally at Level 3 to Form III of the task with the appropriate plot presented first than for Form II. Students were more likely to employ basic data reading in response to Form II than Form III. The less stringent criteria for coding Forms II and III resulted in higher mean scores for these two forms of the task.

Table 4

Examples for Part a) of Forms II and III with Percent in Each Level

Level		Examples for Part a) for Form II	Examples for Part a) for Form III
0	No response or idiosyncratic	"Nothing." "?" **11.8%**	"Don't quite know." "It has 23." **15.38%**
1	Inappropriate interpretation	"Most people move on" "That the 3cm has more" **25.47%**	"Graph 1 is different" "That the majority of families lived in their house for four years" **27.92%**
2	Basic data reading	"That the highest is 4 people for 3 years in the town." "A couple of years missed out."**30.43%**	"That the longest persons family to live in town was 37 years." "3 years is the biggest." **14.53%**
3	Summarises data	"More people have lived in town for 3 years." "The mean = 9.64 (2dp) years. The median = 8 years. The range = 37." **32.92%**	"That most of the population has lived in the town for 0-6 years." "That most people had lived in their town for about 5 years. Also, between 10 and 15 years."**42.17%**

Part b) of Forms II and III in Figure 3 was coded similarly to Part c) of Form I. At Level 0 were "no responses" or responses that misinterpreted the question or reported incorrect differences about the data presented. At Level 1, responses included comments focusing on the aesthetic appearance of the plots, often indicating a personal preference for one over the other without any statistical reasoning. At Level 2, attention to detail and comments about the lay-out or spread and/or the accuracy of the graphs were common. At Level 3 respondents compared and contrasted the exactness of the data on both plots, and/or correctly stated the differences between the scales of the plots. Examples of each coding category for Part b) of Form II and III are given in Table 5. The overall percent of responses at each level is also provided, and as can be seen the percents of responses for both forms of the question for each level are similar. Students often misinterpreted this question or focused mainly on aesthetic appearances of the graph. Paying particular attention to detail and commenting on the accuracy and spread of the lay-out to the detriment of more statistically appropriate ideas was also common. Few students explicitly acknowledged the graphs contained the same data and/or commented on the differences between the scales.

Table 5

Examples for Part b) in Form II and Form III of the Task with Percent in Each Level

Level		Examples for Form II	Examples for Form III
0	Misinterpretation of question or incorrect differences reported	"That less people live in their town in graph 2 than graph 1." **35.09%**	"Graph 2 shows you more into the future by two years." **34.19%**
1	Focus on aesthetic appearance or personal preference of graph	"That graph 2 is a lot harder to read than graph 1." **25.16%**	"Graph 1 is in groups, Graph 2 is spread out." **29.06%**
2	Attention to detail, comments about lay out, spread and accuracy	"Graph 2 was easier to tell how long the majority had lived in the town which was 0-5 because it counted in 5s." **28.57%**	"Graph 1 went up in 5's. Graph 2 missed out numbers in between."**27.35%**
3	Acknowledgement of the same data and/or correctly state the differences between the scales	"There is no difference apart from graph 1 to graph 2 except that graph 2 shows the spaces where as graph 1 doesn't." **11.18%**	"Graph 2 skips the years out that aren't being used. Both graphs show the same information (not a difference)." **9.40%**

Part c) of Forms II and III in Figure 3 was coded in a similar fashion to Part c) of Form I. Again Level 0 responses were those with idiosyncratic reasoning or no reasoning. Level 1 included responses that chose the statistically inappropriate plot, with non-statistical and/or personalised reasoning. At Level 2, respondents chose the statistically appropriate plot but with personalised and/or non-statistical reasoning, or were indifferent in the choice of which plot was better, and stated them as being "both the same" or "neither better" with reasoning reflecting statistical understanding. For Level 3, responses acknowledged the statistically appropriate plot as telling the story better, and provided concise statistical reasoning for the choice. Examples of each coding category for Part c) of Form II and III are given in Table 6. The percent of responses for each level for the two forms of the question are similar. More students responded to Form III in a statistically appropriate manner citing concise and statistical reasoning than for Form II. This outcome was similar to Part c) for Form I. The overall mean scores for Part c), however, were very similar for all three forms. Overall, the modal response was for an inappropriate choice of graph for both forms of the question.

Table 6

Examples for Part c) in Form II and Form III of the Task with Percent in Each Level

Level		Examples for Form II	Examples for Form III
0	Idiosyncratic reasoning or no reasoning	"I think that graph 2 is better because it has more numbers." **21.43%**	"Graph 2." **21.65%**
1	Statistically inappropriate choice with non-statistical or personalised reasoning	"Graph 1 is much better because you can tell a lot easier how many years the people have lived there."**35.09%**	"Graph 2 cause the measurements are better." **37.61%**
2	Statistically appropriate choice with personalised or intuitive reasoning or indifference to choosing	"Neither because they tell the same amount of information." **30.75%**	"I think graph 1 because the answers are more spread out and easy to see."**22.51%**
3	Statistically appropriate choice with concise and statistical reasoning	"Graph 2 because it shows it in half decades and shows the long stop from 25 years to 37 years." "Graph 2, you immediately see there are spaces between the years that families had moved here." **12.73%**	"I think Graph 1 is better. I believe it the graph shows every year and not just the years it needs, it is a lot more informative." "The first is better because it shows how spread out the immigration to town was." **18.23%**

Differences for Grades and Forms

Table 7 contains the means and standard errors for the three grades that participated in Study 1 answering Form I. As can be seen the average level of performance increased with grade ($F_{2,487} = 5.04$, $p < .001$). This reflected increased means for each part for each grade indicating a greater appreciation of the message in the stacked dot plots and of the appropriate display.

Table 7

Means and Standard Errors for Form I of the Task in Study 1

	G5	G7	G9	Total
n	156	162	172	490
Mean	3.76	4.17	4.59	4.19
Standard Error	0.166	0.184	0.200	0.108

Table 8 contains the means and standard errors for grades 5 to 10 on Forms II and III. Overall there was virtually no difference in average performance over the six grades when answering Form II, which presented the stacked dot plot with no gaps for missing values first. When the appropriately scaled stacked dot plot was presented first, in Form III, monotonic and significant improvement was seen in the average performance over the six grades ($F_{5,333} = 10.85$, $p < .001$). In comparing the performance of each grade for Forms II and III of the task, grade 5 students performed better on Form II ($p < .03$) and grade 10 students performed better on Form III ($p < .01$). For all other grades there were no significant differences in performance on Forms II and III. Although there were different grade levels involved in the studies, generally students in grades 5, 7, and 9 did marginally better on the alternative forms than on Form I; however, as noted earlier, this is likely to be a product of the criteria used for Part a) of Forms II and III.

Table 8

Means and Standard Errors for Form II and Form III in Study 2

		G5	G6	G7	G8	G9	G10	TOTAL
Form II	*n*	85	29	61	44	40	52	311
	Mean	4.02	4.66	4.69	4.66	4.60	4.83	4.51
	Standard Error	0.197	0.425	0.262	0.264	0.341	0.338	0.117
Form III	*n*	47	93	46	48	59	43	336
	Mean	3.23	4.02	4.33	4.56	5.20	6.23	4.52
	Standard Error	0.297	0.208	0.342	0.290	0.312	0.360	0.127

For Form II in Study 2 there was a significant difference between male and female participants. This resulted from a significant difference in grade 10 ($p < .03$), with female students scoring higher than male students. No other grades showed gender differences. There were no significant gender differences found in Study 2, Form III. Similarly, in Studies 1 and 3 (Form I) there were no significant gender differences found. Given the number of tests performed, the one observed difference is likely to have occurred by chance.

Change after Instruction

For the students in Study 3 who participated in a series of lessons on chance and data emphasising variation (Watson & Kelly, 2002a; 2002b) the pre and post survey means and standard errors for Form I of the task are given in Table 9. As can be seen the pre-test mean was higher for grade 5 than grade 7 but the grade 7 students showed greater improvement than the other two grades. The improvement for grades 5 and 7 was statistically significant ($p < .01$ and $p < .001$), whereas the improvement for grade 9 was not significant. Although the grade 9 group improved the least, it began with the highest pre-test mean.

Table 9

Means and Standard Errors for Form I in Study 3

		G5	G7	G9	TOTAL
	n	72	79	76	227
Pre	Mean	3.79	3.66	4.61	4.02
	Standard Error	0.259	0.260	0.293	0.158
Post	Mean	4.56	5.13	5.08	4.93
	Standard Error	0.235	0.263	0.319	0.159

Modal Clumps

Of interest, given the observations of Konold et al. (2002) during student interviews of identification of modal clumps, was the frequency with which students in a survey situation would mention features that could be identified with modal clumps. In a survey there could be no prompting to identify such features.

For Form I, Part b) and Form III, Part a), there were two types of responses that could be associated with identifying modal clumps. One type was an implicit reference to the predominance of data in the lower part of the stacked dot plot. The second type was more explicit, usually noting specific numbers on the axis to define a clump. Examples of both types of responses are found in Table 10. For Form I, Part b) in Study 1 implicit responses were given by 34 students (23.9%) who gave Level 2 responses and 14 students (14.1%) who gave Level 3 responses. This represents 9.4% of all responses in Study 1. For Form III, Part a), in Study 2, an implicit response was given by one student (2%) who gave a Level 2 response and 30 students (20.3%) who gave Level 3 responses. This represents 9.2% of all responses in Study 2 for Form III. For Form I, Part a) in Study 1 explicit responses were given by 25 students (17.6%) who gave Level 2 responses and 43 students (50.6%) who gave Level 3 responses. This represents 13.9% of all responses in Study 1. For Form III, Part b), in Study 2, explicit responses were given by 6 students (12%) who gave Level 2 responses and 84 students (56.8%) who gave Level 3 responses. This represents 26.8% of all responses in Study 2 for Form III.

For Part c) of all three forms of the task, appreciation of clumping was found within Level 3 responses that explicitly referred to data being bunched or grouped together in some fashion in the scaled plot. Examples are again given in Table 10. For the three forms of the task the number of students with explicit responses were 9 (10.5%) for Form I in Study 1, 4 (10.3%) for Form II and 13 (20.3%) for Form III in Study 2. This represents about 2% of the overall responses in both studies. Much larger percents of responses for each form of the task focused on the scale and/or the gaps in the data, rather than on the clumps or groups. Typical responses of this type are shown in Tables 3 and 6.

For students who answered the post survey in Study 3, responses could be classified in a similar fashion to above. After instruction the percent of Level 2 responses to Form 1, Part b) that were considered as having an implicit reference to clumps dropped to 11.8%. For Level 3, it was 13.2%. Explicit responses to clumps in Part b) rose to 26.9% for Level 2 but were stable at 48.5% for Level 3. For Part c) reasoning associated with groupings in the appropriate plot accounted for 18.2% of Level 3 answers. Hence there was little indication of change in mention of modal clumps after instruction.

Table 10

Examples of Unsolicited Comments on Modal Clumps

Implicit Form I, Part b); Form III, Part a)	Explicit Form I, Part b); Form III, Part a)	Explicit Form I, II, III, Part c)
"Most people haven't lived there very long."	"That more people lived in town for 6 and under years."	"It shows the families are in groups so it's easier to work out."
"Lots of people came recently."	"0-5 = highest."	"Mostly around the same spot."
"That lots of people have been in town for a short time."	"That they came in 2 groups. May be when someone said it was a great town to live in."	"Graph 1 makes you target a particular group not like Graph 2 where it is spread out heaps."
"Some have stayed longer than others and some only just moved in."	"There are bunches of people who have stayed between 0 and 5 years and 10 and 15 years."	"It has separated the families into 2 groups of 5 years which is easier to interpret."
"I can tell it is a fairly new town for most people especially since only two people have been there for over 20 years."	"Probably a "boom" of population. Populations increase dramatically in a short period of time, then a drought."	"[Graph 2] has a bigger impact as it shows the peaks better than graph 1."

Discussion

The discussion will focus on four aspects of the research: the outcomes observed in terms of understanding over the middle years of schooling, the differences in performance for the different forms of the task, the observations of this study in comparison with the views of Konold and Higgins (2002) and Konold et al. (2002), and the outcomes of this study within the context of the wider goals of statistical literacy within the school curriculum. Finally, the authors will state their subjective views on the nomenclature used to describe stacked dot plots.

Observed Understanding

The levels of response identified for the tasks used in this study show a wide range of student appreciation of the message conveyed by a stacked dot plot. For Form I, response scores ranged from 0 to 9 for every grade in Studies 1 and 3, except for grade 5 on the post survey, which had no scores of zero. The distribution of responses hence is encouraging in that at all three grades some students were able to achieve the highest possible score. The overall percents of responses at each level indicate, however, that there is a long way to go in achieving the goal of appreciation of the message in context. The trend for higher average performance with grade in Study 1 may reflect general ability or exposure to graphing generally, as the stacked dot plot was not part of the curriculum specifically recommended for these grades. The results of Study 3, however, where the intervention lessons focussed primarily on stacked dot plots, reveal the benefits of using stacked dot plots in lessons and of providing an opportunity for students as young as grade 5 to improve their understanding of data display and interpretation by using these representations. The lack of improvement for grade 9 students in Study 3, however, may reflect the results of streaming in the schools involved in the teaching intervention, and the fact that the grade 9 students surveyed were considered by their schools as of average ability. The results of Study 3 should encourage teachers throughout the middle school years to consider the potential of stacked dot plots as tools for learning about graphing and data display.

Different Forms of the Task

In comparison to Form I, for Forms II and III many grades had no zero scores. For Part a) of Forms II and III, this is a reflection of the criteria used for coding and perhaps the motivation of students to attempt the task. Also, for grades 5, 8, 9, and 10 on Form II and grades 5, 7, and 8 on Form III there were no scores of 9. The failure to achieve scores of 9 may represent the presence of two (Parts b) and c)), rather than one (Part c)), summary type question in Forms II and III. As synthesising information is traditionally recognised as more difficult than describing a situation (e.g., Biggs & Collis, 1982), this may account for a drop in frequency of the highest score. The observation of an overall higher average score for Forms II and III than Form I must be reconciled with this observation. It is likely to be due to the scoring criteria for Part a), to the samples involved with some grade 10 students included in Study 2, and to the possibility noted above that the students in Study 2 may have been more highly motivated to engage in the task.

One possible explanation for the difference in pattern of performance over grades 5 to 10 for Forms II and III is related to the lack of appropriateness of the stacked dot plot without gaps, presented first for comment in Form II, and to the other mathematical graphing experiences students build up across grades 5 to 10. It is likely that over the grades 5 to 10 students are exposed to many statistical and algebraic graphs where, within the given limits, scales include all interval values between the limits. When presented with the appropriately scaled stacked dot plot first (Form III), they increasingly with grade identify it as appropriate. This initial reinforcement assists in recognising the deficiencies in the stack dot plot presented second.

For students presented with the stacked dot plot with no gaps first (Form II), they may immediately experience cognitive conflict with their other graphing experiences. This may be more difficult to express and argue than agreement with the plot seen first in Form III for several reasons. First there may be a psychological factor associated with the belief that "a test would not present a bad answer," which must be overcome. The way in which students anticipate events (including tests) may influence the way they react to the situation (Kelly, 1955). An over-dependence on presented information, limited idea exploration, the dominance of the situation, and the need to be right are all obstacles that limit the construction and reconstruction of understanding (Monaghan & Monaghan, 1985). In cases of subtle difference, it is the way in which students' anticipate and construe events that make it more difficult to argue against a suggested perspective than to agree with it. Students answering Form III build up evidence supporting the appropriate plot first, which assists in criticising the second inappropriate plot. In Form II if initial support is offered for the inappropriate plot, this has to be "undone" and replaced in the light of evidence from the second plot.

It would appear that if there were a desire to discriminate across grades with one of Forms II or III of the task, then Form III is preferable, since the overall mean (for all grades) for the two forms is the same. Although it is not possible to draw definitive conclusions based on the samples and means for grades 5, 7, and 9 of this study it might be hypothesised that Form I, due to its use of two more straightforward descriptive parts assists grade 5 students, whereas Form III, due to its structure encouraging more reflection and comparison with the appropriate plot presented first, may assist older students. More research would be needed to support this view.

Comparison with Konold and Higgins

As noted earlier, the task used in this study was based on a case study in Russell et al. (2002) discussed in detail by Konold and Higgins (2002). The context for their discussion was a consideration of beginning efforts by young students to develop their own representations for the data they collected. In moving away from tabular forms and pictographs to more abstract representations, stacks of objects along an axis (line) are a natural progression. A first step is a representation of ordered stacks of blocks, for example, one for each child, with the number of teeth lost represented by the number of blocks. Konold and Higgins then describe a representation of side-by-side stacks of cubes in pairs where one stack represented the number of teeth lost and the other stack represented the number of people who had lost that many teeth. Stacking of case-values and frequencies are more commonly done separately. The plots used in the task in this study are frequency plots but confusion was observed for some students with there being numbers (frequencies of years) represented along the horizontal axis as well. The transition is hence not always easily made.

Other issues at the introductory stage of plot construction include the choice of a range of values in relation to the actual data values to be plotted (e.g., starting at zero if there is no frequency for zero) and whether to aggregate scores in groups (e.g., using a bar for values 0-4 and 5-9). Konold and Higgins (2002) make the following observation:

> As students gain more experience with scaling decisions, they come to see that there is no ideal scale that will make the data appear as they "really" are. Thus it is best to try out several alternative plots and scales and learn what we can from each. When it comes time to summarize results for others, we

> select those representations that do the best job of telling the story sharply and fairly. (p. 180)

In particular the issue of values for zero frequencies for data points is considered important and Konold and Higgins give an extract from a class debate about whether zero should be included when representing frequencies of measured characteristics. As well they comment on the usefulness of the appropriately scaled plot (Graph 2 in Figure 2) for showing

> two distinct clusters, with one group of families having lived in town between 0 and 6 years, and the other between 10 and 14 years... Seeing these two separate clumps may raise interesting questions about factors affecting town growth. Indeed, an economic recession had hit the area beginning about 1989, which more recently had abated. (p. 182)

The other plot (Graph 1 in Figure 2) makes identifying of clusters very difficult. This aspect of comparing the two stacked dot plots was noted by a very small percent of students overall in this research. Without any kind of prompting students tended to focus on scale and gaps rather than clusters in their responses. Although it can be argued that in some sense the comments on gaps are complementary to the comments on clusters, it is of interest that more often students focused on gaps. For the task in this study it may be that the alternative plot, which did indeed represent the same frequencies in groups but without gaps, fostered the "gap" responses. Another form of questioning would be required to ascertain students' relative propensities to focus on clumps or gaps, or even the outlier. Overall the issues discussed by Konold and Higgins and Konold et al. (2002) are more wide-ranging than considered in the questions asked in this study but they point to the context within which this type of activity should be used in the classroom.

Statistical Literacy and the Classroom

The task used in this study satisfies the two components of statistical literacy as set out by Gal (2002):

> (a) people's ability to *interpret and critically evaluate* statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to *discuss or communicate* their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions. (pp. 2-3)

The task in this study asks students to interpret the meaning of the plot and evaluate which form is better for the purpose of telling the story in the context and to write coherently about their observations. The increasing use of graphs in the media (e.g., Rothstein, 2001) means that approaching this topic while students are still in school is likely to enhance their ability to transfer their statistical literacy skills to the adult environment. A way of assessing students' progress to the goal of statistical literacy is provided by Watson (1997) in a three-tiered hierarchy. Tier 1 emphasises an understanding of basic statistical terminology. Tier 2 covers applying the terminology when it is found in social contexts. Tier 3 relates to the critical thinking required to question claims that are made without appropriate statistical foundation. In Form I, Parts a) and b) address the aspects of the first two tiers of the hierarchy, basic interpretation and placing that understanding in context. Part c) asks for a critical assessment of the two plots, consistent with the third tier with regard to choosing a statistically appropriate representation. In Forms II and III, Part a) again satisfies Tiers 1 and 2, whereas Part b) provides scaffolding for Part c), which again addresses Tier 3. That up to about one third of students have difficulty engaging the task at a level recognising what story the plot is telling (i.e., cannot reach Tier 2) is a concern, as are the low percents (between about 9 and 18 percent) who can communicate successfully their reasons for selecting the appropriate plot as telling the story better (Tier 3).

Acknowledging that the stacked dot plot is one of the simplest representations for students to create (Bereska, et al., 1998) and hence to interpret, it appears that teachers and curriculum planners need to be explicit in addressing these issues related to statistical literacy. Russell and Corwin (1989) and Bereska et al. (1998, 1999) provide excellent starting points at the primary school level, as do Landwehr and Watkins (1986) at the secondary level and Dossey et al. (2002) for preservice secondary mathematics teachers. Tasks such as those used in this study are also adaptable for the classroom either in the format used here or based on students collecting and representing their own data. If no one uses the form of Graph 1 in Figure 2, it can be introduced to provide cognitive conflict to stimulate student debate. Graphing is an important component of statistical literacy that is necessary for understanding in many parts of the school curriculum, including science, social science, health, and technology, as it will be when students leave school.

Nomenclature

Having acknowledged in the introduction to this paper the many different terms that have been used to refer to the type of graph employed in this study, the authors advocate the use of the term *stacked dot plot*. Although the term line plot appears to be the most popular in materials for the primary and middle school levels (Bereska et al., 1998, 1999; NCTM, 2000), there is still much confusion expressed from statisticians looking down from the higher echelons. Their disdain for the term does nothing to help communication among the different levels of education. The use of the term dot plot on its own also creates confusion from several perspectives. Wilkinson (1999) distinguishes asymmetric dot plots (like the plots in this study) from symmetric dot plots that are suspended above the base line symmetrically about an imaginary central line, and also refers to a histodot plot where dots are placed in bars. Although he refers to *stacks of dots*, he does not use the term stacked dot plot for his asymmetric dot plot of individual values. Cleveland (1993) uses multiway dot plots to display relationships among several variables in such a way that the dots are not stacked but dispersed to tell the story in the data.

It is our view that the term stacked dot plot is illustrative in the words used, all of which are comprehensible to primary aged children. Although "dot" can be confusing if Xs are used instead of dots, the generic representation seems less confusing than the mix-up between the "base line" and the "line connecting points" (see Figure 1) that occurs when the term line plot is used. Stacking Xs (or dots) on the line is natural for beginners and the authors agree with Konold (personal communication, 26 March, 2002) that the phrase *stacked dot plot* is the most sensible at the school level.

Whatever the term used, these plots provide a straightforward means of summarising small data sets to gain an initial feeling for the centre, spread, and shape of the distribution involved. Their introduction to all classrooms should not be delayed until the argument about nomenclature is resolved.

Acknowledgements

This project was funded by an Australian Research Council grant (No. 00000716) and by the University of Tasmania Institutional Research Grant Scheme (No. W0011912). Annaliese Caney coded Forms II and III of the questionnaire in Study 2.

References

Bereska, C., Bolster, C. H., Bolster, L. C., & Schaeffer, R. (1998). Exploring statistics in the elementary grades. Book One (K-6). White Plains, NY: Dale Seymour.

Bereska, C., Bolster, C. H., Bolster, L. C., & Schaeffer, R. (1999). Exploring statistics in the elementary grades. Book Two (grades 4-8). White Plains, NY: Dale Seymour.

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

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). Graphical methods for data analysis. Pacific Grove, CA: Wadsworth & Brooks/Cole Publishing Company.

Cleveland, W. S. (1993). Visualizing data. Summit, NJ: Hobart Press.

Collis, K. F., & Biggs, J. B. (1991). Developmental determinants of qualitative aspects of school learning. In G. Evans (Ed.), Learning and teaching cognitive skills.Hawthorn, VIC: The Australian Council for Education Research.

Dossey, J. A., McCrone, S., Giordano, F. R., & Weir, M. D. (2002). Mathematics methods and modeling for today's mathematics classroom: A contemporary approach to teaching grades 7-12. Pacific Grove, CA: Brooks/Cole.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. International Statistical Review, 70, 1-51.

Kelly, G. A. (1955). The psychology of personal constructs: A theory of personality. Vol. 1. New York: W. W. Norton & Company Inc.

Konold, C., & Higgins, T. L. (2002). Working with data: Highlights related to research. In S. J. Russell, D. Schifter, & V. Bastable (Eds.) Developing mathematical ideas: Collecting, representing, and analyzing data. Parsippany, NJ: Dale Seymour Publications.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), CD of the Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Landwehr, J. L., & Watkins, A. E. (1986). Exploring data: Quantitative literacy series. Palo Alto, CA: Dale Seymour Publications.

Lappen, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (1997). Comparing and scaling: Ratio, proportion, and percent. Palo Alto, CA: Dale Seymour Publications.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. (2nd ed.). Thousand Oaks, CA: Sage Publications.

Monaghan, L., & Monaghan, R. (1985). Wondering constructions of learning: Lateral and vertical styles in children. In F. Epting & A. W. Landfield (Eds.), Anticipating personal construct psychology (pp. 128-139). Lincoln: University of Nebraska Press.

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

Rothstein, R. (2001, November 28). Statistics, a tool for life, is getting short shrift. The New York Times. Retrieved May 10, 2002 from http://www.nytimes.com/2001/11/28/national/portraits/28LESS.html.

Russell, S. J., & Corwin, R. B. (1989). Used numbers: The shape of data. Palo Alto, CA: Dale Seymour Publications.

Russell, S. J., Schifter, D., & Bastable, V. (2002). Developing mathematical ideas: Collecting, representing, and analyzing data. Parsippany, NJ: Dale Seymour Publications.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal & J. B. Garfield, The Assessment Challenge in Statistics Education (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

Watson, J. M., & Kelly, B. A. (2002a). Grade 5 students' appreciation of variation. Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education. Norwich, UK: PME.

Watson, J. M., & Kelly, B. A. (2002b). Variation as part of chance and data in grades 7 and 9. Proceedings of the 25th Conference of the Mathematics Education Research Group of Australasia: Mathematics in the South Pacific, Auckland. Sydney: MERGA.

Wilkinson, L. (1999). Dot plots. The American Statistician, 53, 276-281.