

Interpreting Graphs: Students Developing an Understanding of Covariation

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

Students' development of an understanding of covariation was the focus of a research project that investigated the way in which 12 Year 5/6 students engaged with the learning environment afforded by the graphing software, *TinkerPlots*. Using data generated from individual interviews, the results demonstrate that upper primary students use their understanding of covariation to draw conclusions about trends in data evidenced in the graphs they created as well as their knowledge of the context. Implications for the Australian Curriculum are considered.

Covariation is used to explore the relationship between two variables and is recognised as an important aspect of statistical reasoning. It is determined by examining graphs of bivariate data that are represented in scatterplots. Scatterplots are an economical way of organising large amounts of data that provide information about two variables that are not necessarily dependent on each other and show the correspondence of the ordination of each variable (Moritz, 2004, p. 228). They also assist in the identification of the relationship between two variables and variation from that relationship as well as being particularly useful for identifying clusters of points and outliers in a distribution of bivariate data.

The characteristics of graphs, such as the mode, scale of an axis, or the variation in the spread of the data can be extracted directly from a graph (Kosslyn, 1989) or from calculations performed by graphing software, such as *TinkerPlots: Dynamic Data Exploration* (Konold & Miller, 2005). Another factor that influences conclusions that are made about graphs is the context of the data (Watson, 2006). An understanding of the context and the nature of the variables of interest may be gained from personal experiences of the context, from information about the data, or from the details embedded in the characteristics of the graphs created from the data. Collectively, the elements of graphs are resources that provide a link between the two-dimensional representation of data and the context of real world measurement situations. Graphing software such as *TinkerPlots*, provides access to the characteristics of graphs and allows data analysis activities to hone in on the story graphs have to tell and focus on how graphs can be used by students to evidence their thinking about the data. To date, few studies have addressed the way in which students construct and evaluate graphs within technological learning environments (Shaughnessy, 2007).

Interpreting Covariation in Graphs

Cartesian graphs, such as scatterplots, depict covariation of two sets of measurements that vary along numerical scales in a two-dimensional space. Scatterplots, in particular, are characterised by data points that correspond to the measures of two variables designated at the same time. Each data point also corresponds to one unit of analysis between the two variables. The values of the two variables may be said to involve some form of relationship, association, function, dependency, or correspondence (Moritz, 2004). In this report, the emphasis for covariation is on the relationships between two variables, which are interpreted as general trends that show the variation of the two variables due to the ordination of the values along each axis of a graph with numerical scales.

Although the use of scatterplots to determine a trend in data has been explored at the middle school level, few studies have focused on primary students. One exception is a study of Year 3, 5, 7, and 9 students that looked specifically at speculative data generation that involved students translating a scatterplot into a verbal statement, interpreting numerical data, reading values from a graph, and interpolating data (Moritz, 2004). Moritz found that students tended to focus on individual data points rather than look globally for the trend in the data and often only considered one of the variables in isolation. Ben-Zvi and Arcavi (2001) also found that Year 7 students focused on particular points of a graph. They noticed, however, that this did not always constrain the students and “served as a basis upon which the students started to see globally” (p. 62).

In 2003, Cobb, McClain, and Gravemeijer set up a classroom experiment that explored 29 Grade 8 students’ understanding of bivariate data, with a particular emphasis on statistical covariation. The part of the investigation that explored the way in which the students used scatterplots to analyse data found that focusing on the density and the shape of data was crucial for the students to interpret scatterplots. They also noted that students “typically reduced scatterplots to lines that signified fixed relationships of covariation rather than conjectured relationships about which the data were distributed” (p. 75).

The Study

The study, from which this paper was taken, employed an educational design research methodology (Akker, Gravemeijer, McKenney, & Nieveen, 2006) within a pragmatist paradigm (Mackenzie & Knipe, 2006). This paradigm was adopted as it positioned the study as being scientifically-based research that could potentially inform and develop evidence-based practice in education. It facilitated the development of a systematic study, whose purpose was to capture the way in which *TinkerPlots* contributed to, supported, and influenced the students’ development of understanding of covariation as they created and interpreted graphs.

The study involved the development of a learning sequence and its subsequent implementation with 12 Year 5/6 students. Over a period of six weeks (45 minute sessions, twice a week), the students worked in pairs to complete a series of assigned tasks that included collecting data, constructing a range of graphs, comparing data sets of varying size, and identifying trends and outliers using *TinkerPlots*. At the end of the learning sequence, each student was interviewed individually using an interview protocol set up in *TinkerPlots*. The interview protocol required the students to explore a data set by constructing graphs and making conclusions and informal inferences about the data seen in the graphs. As the students worked at the computer, they were asked questions that required them to describe their graphs and explain how they made their conclusions about the data from their graphs. The interview protocol was designed to be a culminating performance of understanding. As such, it was open-ended and required the students to draw on and demonstrate their skills and understandings developed throughout the sequence of learning experiences. The data set used in the interview protocol was extracted from the Census@Schools data base (www.abs.gov.au). It was a random sample of 200 Year 5/6 students from across Australia generated from the Census@Schools website. The attributes in the data set were: gender, foot length, height, and belly button height.

As the students worked through the interview protocol with *TinkerPlots*, the authoring software package with an on-screen recording utility, *Adobe Captivate 3* (Adobe Systems Incorporated., 2007), was used to create a video of the students’ activity as they worked on the computer. The *Captivate* files generated showed the screen of the computer as seen from the user’s perspective. The files recorded all movements of the cursor as the students

worked at the computer and provided a chronological record of how each student accessed the features of *TinkerPlots*. *Captivate* also provided audio data of the conversations the students had with the researcher during the interview. The audio data from the files provided evidence of what the students said and the on-screen video data provided evidence of what the students did.

The audio data from the interview videos generated by *Captivate* were transcribed verbatim. The audio transcripts were then synchronised with the on-screen capture video data generated by *Captivate* and descriptors of the students' actions at the computer were added to the transcripts. In the first part of the analysis process, the data excerpts from the audio transcripts and the descriptors of the students' actions were coded according to the four key elements of the *Graphing in EDA Software Environments* framework (Figure 1). After this initial categorisation, the data that were coded for the key elements Understanding Data and Thinking about Data were analysed to determine the way in which the students identified, explained, and interpreted trends and relationships between two attributes in graphs. The aim was to determine the ways in which students develop an understanding of covariation. This paper reports on the results from the second part of the data analysis process.

Categories		Key Elements
<i>Generic Knowledge</i> Speaking the language of data and graphs. Recognising the characteristics of data and graphs. Understanding how to use the features of software and technology	<i>Being creative with data</i>	Reducing data to graphical representations. Summarising data. Constructing different forms of graphs. Describing data from graphs.
	<i>Understanding data</i>	Making sense of data and graphs. Understanding the relationship among tables, graphs, and data. Identifying the messages from the data. Answering questions about the data. Recognising appropriate use of different forms of graphs.
	<i>Thinking about data</i>	Asking questions about the data. Recognising the limitations of the data. Interpreting data. Making causal inferences based on the data. Looking for possible causes of variation. Looking for relationships among variables in the data.

Figure 1. Framework for Graphing in EDA Software Environments (Fitzallen, 2006, p. 206).

Analysis

Kosslyn (1989) contends that to analyse a graph it is necessary to understand the interrelated connections among the constituents of a graph. He asserts that understanding the interrelated connections fosters the interpretation of graphs on three levels. First, the individual elements and their organisation can be described. Second, understanding of the display can be determined by looking at the relationship between the elements of the graph. Lastly, the analysis can extend to the interpretation of the symbols and lines that goes beyond the literal reading of the information. Influenced by the work of Kosslyn, Curcio (1989) considered school students' interpretation of graphs from three perspectives: reading data directly, reading "between" the data, and reading "beyond" the data. These phrases reflect to some extent the increasing demands of the levels suggested by Kosslyn (1989). In 2007, Shaughnessy added to the work of Curcio by suggesting the further need to read "behind" the data. The additional dimension suggested by Shaughnessy (2007) together

with the reading “beyond” the data perspective offered by Curcio, reflects the complexity of graph interpretation at the third level suggested by Kosslyn.

The three levels of graph interpretation highlighted by Kosslyn (1989) and others were used to devise a framework for analysing the students’ responses about covariation and the relationship seen in the data from the graphs they created. The framework devised was based on the Structure of Observed Learning Outcomes (SOLO) model developed by Biggs and Collis (1982). It has three levels that describe the increasing complexity of thinking required when interpreting graphs. At the first level, termed uni-structural, responses employ single elements to describe the relationship seen in the data. These are isolated statements that are not linked together to draw conclusions. This aligns with level 1 of the Kosslyn framework. At the second level, termed multi-structural, responses employ multiple elements to describe the relationship seen in the data. These are offered in the same excerpt of conversation as a string of ideas that are supporting evidence about the relationship seen but are not presented in an interrelated fashion. This aligns with the second level of the Kosslyn framework. At the third level, relational, responses employ multiple elements to describe and justify the relationships seen in the data and make explicit how the elements are interrelated, including elements beyond the literal reading of the data, such as the context. This aligns with the third level of the Kosslyn framework.

Results: Students’ Understanding of Covariation

The students’ statements, descriptions, and justifications about the relationship seen in a graph made during the completion of the interview protocol were analysed to determine the complexity of the responses according the SOLO framework devised reflecting Kosslyn (1989). Of the 12 students interviewed, six students’ responses were uni-structural, three students’ responses were multi-structural, and three students’ responses were relational. Each of the students’ response at each level is described in turn.

Uni-structural Thinking

The results from this study suggest that upper primary students begin with an intuitive appreciation of covariation that is based on their knowledge of and experience with the context of the data. When the students in the study who were operating at this uni-structural level were asked to describe how a relationship identified was evidenced in the graphs they constructed, they were able to offer information from the graphs to support their statements but often their justifications and examples were offered piece-meal without making any connections between the different constituent parts of the graphs and without making connections between the graphs and their knowledge of the context. The students could also identify a trend in a graph and use it to determine the relationship between two attributes but their explanations about how they used the trend to determine the relationship were incomplete. Often their statements involved a declaration that there was or was not a trend evident but little or no justification or reasoning was offered to explain how they made the judgement.

The use of isolated statements was evident when Jake suggested “Umm ... The difference between the tallest, the highest one and the mode is ... 17, I think” as he extracted information from a graph. When making this statement he did not make any connections between this information and the generalised conclusion he made previously for the same graph, “Hmm ... the ... the foot’s not really getting any bigger. It’s staying kind of the same.” Most of the time, his judgements were based on the comparison of individual

values and ranges of data in a graph and he scrutinised the two variables in a scatterplot separately.

Natasha based her judgements about the relationship between two attributes on her knowledge of the context. For each of the covariation graphs she created, she made generalised statements like, “Yeah. ‘Cos the taller you are, like ...there’s usually got like a higher... higher belly button height, if that makes sense?” She was also able to make conjectures about the relationship but was limited to considering one data point at a time and did not make any collective statements to suggest her thinking was shifting beyond being uni-structural. Rory also worked in a similar manner, reporting, “Yeah that’s the height and, well the big people have the big foot, sort of.” Like Natasha, the evidence he offered to support his statement focused on the value of specific data points.

Johnty’s statements about the covariation graphs were uni-structural and for the most part, relied on the comparison of individual data values or were based on the physical location of the data in the graphs. This was evident when he said a relationship existed between the attributes belly button height and height because the dots in the graph were “going up.” He also relied on his knowledge of the context to make his conjectures when he said “The taller you are, usually, the higher your belly button is,” without referring to the graph or the data. In contrast, Kimberley made multiple statements about the graphs that showed she could extract a variety of information from the graphs. This showed she had an appreciation that there was a relationship between the two attributes in a graph but her isolated statements about the data were presented as a list of individual characteristics without connections being made between the statements about the trend or the multiple aspects identified. Her statements included, “Mmm ... well the higher the person is the longer the foot length ... seems” and “this one here’s the highest, and with about 185 but he’s only got a length of 25 cm for the foot.” Kimberley did not attempt to combine her statements to make conclusions.

Like the other students’ responses at the uni-structural level, Natalie identified appropriately from her knowledge of the context that “the smaller you are the smaller your foot length is.” There were, however, instances when Natalie’s responses were beginning to merge into multi-structural. This occurred when she was making a decision about an outlier. She said, “Umm, it’s an outlier because of, like, you wouldn’t think it would be 85 because she’s 124 and I’m taller than her and mine is only 65.” Natalie did not, however, make enough statements at the multi-structural level for her thinking to be considered multi-structural.

Multi-structural Thinking

The students operating at the next level of understanding, multi-structural used multiple constituent parts of their graphs to explain their thinking and to describe the covariation identified. Their responses, however, did not show they had an appreciation of the way in which different elements of the graphs were related and how they could be used to support conjectures and conclusions. They were able to identify regions of the graphs that varied from the trend identified but were unable to explain how the variation informed their overall conclusions. Blaire, for example, when asked about the relationship seen in a covariation graph displaying the attributes belly button height and height, moved immediately to say, “Ok, as your ... of course you would already know that as you get taller your belly button height would get higher.” She then went on to declare that the mean belly button height would be in the middle of a cluster of data in the graph. Although adding the mean to the graph using the function in TinkerPlots corroborated her assumption about the mean, she

could not go further to use the mean to justify her generalised statements about the relationship shown in the graph.

Jessica's also used her knowledge of the context as a starting point for evaluating her graphs, "That yeah, as you grow, your foot length usually grows too." When asked to describe how the graph showed this she went on to say, "Well like there ... the largest person, their foot length is only about 25 cm whereas this person here, they have the largest foot length but they're not the tallest." Jessica was making sense of the data by considering different aspects of the graph but she did not tie the information together successfully to make responses that were relational. Like Blaire and Jessica, Shaun identified variation within the graphs that did not fit with his generalised conclusions but he was unable to make confident statements about the relationship in the graphs. In one exchange with the teacher, Shaun commented on the relationship evident in a graph by saying, "the clump [in the graph] shows most of the people, the higher the foot they have, the higher the belly button to the floor." He also identified specific data points that were outside the main clump and looked at the values of the attributes for those data points. He did not, however, indicate how the additional information influenced his generalised statement.

Relational Thinking

The students at the relational level had developed more sophisticated ways of describing the relationship between two attributes. They did this from both global and local perspectives. They were able to identify a trend in a graph, identify the variation in the graph that did not meet their expectations for a relationship to exist, and use multiple constituent parts of graphs to support their reasoning. They interlaced their explanations with information about specific data points, the variation, and the trend evident, as well as making connections between the information in the graphs and their knowledge of the context to justify their conclusions. James put these ideas together successfully when he stated "that well, like ... when they were getting taller their belly button height was getting bigger." He then went on to explain details in the graph, "And like from around here. I think them ones are like taller, but their belly buttons are only 50 and 60. They'd have to be very short, like to have a belly button that small. Like these ones are all what you'd expect from like, that age. Whereas, like these ones out here ... like aren't."

Mitchell was also able to relate different pieces of information to each other to make his conclusions. This was evident when he said "Yeah I'd say, well, yeah, there's a couple of people that are off it a bit more [referring to the trend identified in the previous graph] but it's pretty much the same as foot length and height. The taller your belly button is the higher you are." Mitchell's ability to make connections to other graphs displaying different attributes supports further that his thinking was relational. William's explanations started with descriptions of individual data points to support his statements about the relationship between two attributes before he moved on to thinking about the data from a broader perspective. For example, he said, "It tells us that that one's the biggest now and it's also the biggest foot length." As he worked, he extended his explanation to include descriptions of regions of data that he considered to be following a trend and then combined this with descriptions of particular data points within the graph that did not fit the trend.

Discussion and Implications for the Curriculum

It is generally assumed that young children's ability to evaluate relationships in data and reflect on supporting evidence is undeveloped (Moritz, 2004). In his study, Moritz found otherwise. The evidence in this study suggests that students in the upper primary years are

able to evaluate the relationship between two attributes and develop an understanding of covariation. At the beginning level of understanding, they rely on their knowledge of the context of the data or individual elements of graphs to make their conclusions. At the next level they start to consider multiple characteristics of graphs but do not always elaborate on the connections among the characteristics or between the graphs and their knowledge of the context. This increases to having the ability to bring together the elements of the graphs and their knowledge of the context effectively at the highest level. The students operating at the highest level of the thinking for covariation were able to make the connections among the constituent parts of the graph, the context of the data, and the message in the data characterised by the trend. These students were able to express their thinking verbally and talked about their graphs from both generalised and local perspectives. It is therefore, imperative that teaching and learning activities provide the opportunity for students to develop strategies for moving back and forth between thinking about what can be seen in the data from the generalised perspective of expressing the relationship between two attributes as the trend seen in the data to the more localised perspective of using the spread of a cluster of data or outliers in a graph to describe how the thinking about the relationship was formulated. This study demonstrated how *TinkerPlots* provided a flexible learning environment that fostered and supported the development of such flexible thinking strategies. Considering that students in this study are able to describe covariation and generalise about a trend evident in the data it may be pertinent to bring the initial introduction of scatterplots into the upper primary years of schooling. This would provide students with the opportunity to establish the notion of covariation, how it is characterised in graphs, and how to reason about covariation well before the introduction of the statistical analysis of correlation and the translation of a trend into an algebraic expression in Year 10.

Development of covariation skills and the use of scatterplots have the potential to facilitate learning outcomes in the Australian Curriculum (ACARA, 2011). In History, for example, when looking at change over time and when exploring the notion of cause and consequence. Although the History curriculum is not explicit about the interpretation of graphs in their outcome statements, it would be beneficial for the History curriculum to elaborate on where graph creation and graph interpretation skills can contribute to the development of outcomes for the History Inquiry Skills strand. Taking into account that the Year 5/6 students in this study demonstrated the ability to describe and generalise about the relationship between two variables, the use of scatterplots to interpret census data to construct arguments about immigration in Year 6, say, would be beneficial. This would lay the foundation for being able to identify trends and draw conclusions about historical events over time, as required in Year 9 and Year 10.

The Science curriculum recommends that graphs are used to make conclusions, interpret data and make inferences; however, the application of graphs for these purposes in the Science curriculum does not align with when the fundamentals of graph creation and graph interpretation are introduced in the Mathematics curriculum. The Science curriculum suggests that students in Year 2 use column graphs yet they do not appear in the Mathematics curriculum until Year 3. Another issue arises in Year 6 when students are expected to compare data and make predictions as well as interpret trends in data. Bearing in mind the Mathematics curriculum does not address the development of the graphing skills required in Science until the secondary years of schooling, it would be worth rethinking where to place the learning about graphing and data analysis in the Mathematics curriculum and its application in other curricula, such as History and Science. Aligning complementary curricula would allow the Mathematics curriculum to coincide with the demands of other curricula, thereby fostering the connections between different learning areas and addressing

the need for all curriculum areas to embrace the cross-curriculum priorities, detailed in the Australian Curriculum (ACARA, 2011).

Acknowledgements

This research is funded by an APA(I) Scholarship associated with an ARC Linkage Project LP0560543 and the industry partner, Department of Education, Tasmania. Thanks to Prof Jane Watson for reading the initial draft and providing valuable feedback.

References

- Adobe Systems Incorporated. (2007). *Adobe Captivate 3*. [Computer software] San Jose, CA: Adobe Systems Incorporated.
- Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.) (2006). *Educational design research*. New York: Routledge.
- Australian Curriculum, Assessment and Reporting Authority (2011). *The Australian Curriculum*. Retrieved from <http://www.australiancurriculum.edu.au>
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35-65.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning. The SOLO taxonomy*. New York: Academic Press.
- Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1-78.
- Curcio, F. R. (1989). *Developing graph comprehension*. Reston, VA: National Council of Teachers of Mathematics.
- Fitzallen, N. (2006). A model of students' statistical thinking and reasoning about graphs in an ICT environment. In P. Grootenboer, R. Zevenbergen & M. Chinnappan (Eds.), *Identities, Cultures and Learning Spaces* (Proceedings of 29th annual conference of the Mathematics Education Research Group of Australasia, Canberra, pp. 203-210). Sydney: MERGA.
- Konold, C., & Miller, C. D. (2005). *Tinkerplots: Dynamic data exploration*. [Computer software] Emeryville, CA: Key Curriculum Press.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185-226.
- Mackenzie, N. & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues In Educational Research*, 16(2), 193-205. Retrieved from <http://www.iier.org.au/iier16/mackenzie.html>
- Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi, & J. Garfield, (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 227-255). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester, Jr. (Ed.), *Second handbook on research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: Information Age Publishing.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.