



Investigating learners' meta-representational competencies when constructing bar graphs

Author:Michael Mhlolo¹**Affiliation:**

¹Faculty of Humanities
Postgraduate Studies, Central
University of Technology,
South Africa

Correspondence to:

Michael Mhlolo

Email:

mikemhlolo@yahoo.com

Postal address:

Private Bag X20539,
Bloemfontein 9300,
South Africa

Dates:

Received: 22 Feb. 2015

Accepted: 09 June 2015

Published: 30 June 2015

How to cite this article:

Mhlolo, M. (2015).
Investigating learners'
meta-representational
competencies when
constructing bar graphs.
Pythagoras, 36(1), Art. #259,
10 pages. <http://dx.doi.org/10.4102/pythagoras.v36i1.259>

Copyright:

© 2015. The Authors.
Licensee: AOSIS
OpenJournals. This work is
licensed under the Creative
Commons Attribution
License.

Read online:

Scan this QR
code with your
smart phone or
mobile device
to read online.

Current views in the teaching and learning of data handling suggest that learners should create graphs of data they collect themselves and not just use textbook data. It is presumed real-world data creates an ideal environment for learners to tap from their pool of stored knowledge and demonstrate their meta-representational competences. Although prior knowledge is acknowledged as a critical resource out of which expertise is constructed, empirical evidence shows that new levels of mathematical thinking do not always build logically and consistently on previous experience. This suggests that researchers should analyse this resource in more detail in order to understand where prior knowledge could be supportive and where it could be problematic in the process of learning. This article analyses Grade 11 learners' meta-representational competences when constructing bar graphs. The basic premise was that by examining the process of graph construction and how learners respond to a variety of stages thereof, it was possible to create a description of a graphical frame or a knowledge representation structure that was stored in the learner's memory. Errors could then be described and explained in terms of the inadequacies of the frame, that is: 'Is the learner making good use of the stored prior knowledge?' A total of 43 learners were observed over a week in a classroom environment whilst they attempted to draw graphs for data they had collected for a mathematics project. Four units of analysis are used to focus on how learners created a frequency table, axes, bars and the overall representativeness of the graph *vis-à-vis* the data. Results show that learners had an inadequate graphical frame as they drew a graph that had elements of a value bar graph, distribution bar graph and a histogram all representing the same data set. This inability to distinguish between these graphs and the types of data they represent implies that learners were likely to face difficulties with measures of centre and variability which are interpreted differently across these three graphs but are foundational in all statistical thinking.

Introduction

Traditionally instructional focus in the statistics classroom has been on learners' construction of various graphs with the instruction being didactic in nature but with little attention being given to the analysis of reasons why the graphs were constructed that way in the first place (Friel, Curcio & Bright, 2001). Similar concerns have been expressed by diSessa, Hammer, Sherin and Kolpakowski (1991, p. 157), who have suggested:

One of the difficulties with conventional instruction ... is that students' meta-knowledge is often not engaged, and so they come to know 'how to graph' without understanding what graphs are for or why the conventions make sense.

Watson and Fitzallen (2010) suggest that little is likely to be achieved by providing a collection of data (found in the textbooks) and having children practise drawing graphs in isolation. A recommendation that is consistent with current views of 'data handling' that goes beyond 'statistics' is put forth by Shah and Hoeffner (2002), who suggest that research on learners' abilities to construct graphs, and how this relates to their ability to comprehend graphs, was particularly relevant for project-based activities in which learners create graphs of data that they collect for themselves. Due to the fact that collected data are grounded in real-world contexts, diSessa (2004) argues that an ideal environment is usually created for learners to demonstrate their meta-representational competence. Such competence includes learners' abilities to invent or design a variety of new representations, explain their creations, understand the role they play and critique and compare the adequacy of such representations. Learners' meta-representational competence is the very resource out of which expertise is constructed (diSessa & Sherin, 2000) and a number of researchers have used other terms such as phenomenological primitives (p-prims) (diSessa, 1993, 2004), cues (Davis, 1984) or 'met before's' (Tall, 2008) in support of the existence of such a pool of knowledge.

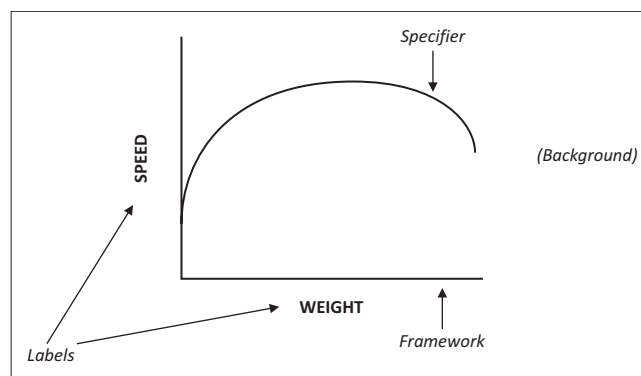


Although previously activated knowledge structures (diSessa, 1993) are acknowledged as critical resources, Tall (2008) cautions that it should not be taken for granted that new levels of mathematical thinking are necessarily built logically and consistently on previous experience. Empirical evidence has shown that the existence of prior knowledge can also lead to negative outcomes in the form of ‘misconceptions’ (English, 2012). Given this dichotomous nature of prior knowledge, diSessa and Sherin (2000) suggest that we should understand this resource in more detail for its theoretical and practical import in learning. We should raise questions about the nature and content of these intuitive ideas, where they come from and how they are involved, both productively and unproductively, in learning. These are the questions that steered this analysis of Grade 11 learners’ instructional activities during the process of constructing bar graphs. The learners worked with data that they had collected for themselves for a mathematics project that was part of their curriculum requirement. The article aims more specifically to tease out evidence of the knowledge representation structures that were stored in the learners’ memory and the extent to which this pool of knowledge was (in)adequate as a resource for bar graph construction.

Basic-level constituents of a graph

Given this objective, it is doubtful whether one could discuss adequacy, productivity or effectiveness in graph construction without making references to conventions that guide us in validating our concept of adequate, truth, correctness and accuracy in such mathematical activities. With this in mind it seems appropriate to develop an understanding of the way graphs are structured to appreciate the way in which they communicate information. In doing so I acknowledge Watson and Fitzallen (2010), who point out that due to the more recent emergence of the field of statistics there is more flexibility on what the conventions should be, unlike algebra and other areas of mathematics where conventions are more fixed.

Despite this variability in nomenclature and conventions, especially in statistical graphs, researchers warn that writing realistic assessment items and resources to mark them would not be easy if there was no movement towards convergence on conventions (Kosslyn, 1989; Shah & Hoeffner, 2002; Watson & Fitzallen, 2010). Consistent with this need to move towards convergence on conventions, this article borrows from Kosslyn (1989) who suggests a schema for the analysis of graphs that can be used to communicate information clearly and concisely. Kosslyn argues that even though there are many types of graphs they are all made up of the same basic-level constituents. The elements include the ‘background’, the ‘framework’, the ‘specifier’ and the ‘labels’ (Kosslyn, 1989, p. 188). Figure 1 illustrates the basic-level constituents of a typical graph.



Source: Kosslyn, S.M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–226. <http://dx.doi.org/10.1002/acp.2350030302>

FIGURE 1: The basic-level constituent parts of a graph.

The background is the pattern over which the other component parts of a graph are presented. In most instances the background is blank as it is not necessary to include a pattern or picture. The framework represents the kinds of entities being related, in this case weight on the x -axis and speed on the y -axis. The specifier conveys specific information about the entities represented by the framework by mapping parts of the framework (in this example weight) to other parts of the framework (speed). The specifier may be a point, line or bar and is often based on a pair of values (x and y values). The labels of a graph are an interpretation of a line or region. They may be letters, words or pictures that provide information about the framework or the specifier. To analyse graphs it is necessary to understand the interrelated connections amongst these constituents of a graph. So how do these basic-level constituents help us to distinguish between the different types of bar-like graphs?

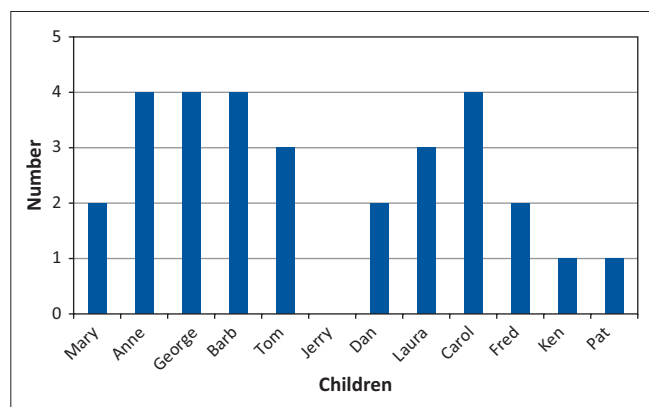
The constituent parts of bar-like graphs

Although there is variability in naming these bar-like graphs in this article I adopt terminology used by Cooper and Shore (2010) as well as Watson and Fitzallen (2010). The decision was guided by what I viewed as (1) the consistency with which their work builds on Kosslyn’s (1989) work, (2) their long-standing history of contribution to making sense with graphs, (3) clarity in the way they exemplified the links between these graphs and (4) the need to maintain consistency in the discussion. Watson and Fitzallen (2010) posit that bar-like representations are of three major types (value bar graphs, distribution bar graphs and histograms), which are presented as historically developing from one into the other in that order. This article does not intend to dwell much on the historical development of these graphs but suffice it to say that, especially at primary and secondary school level, these bar-like representations are often simply referred to as bar graphs, so that their distinction is unclear. This is despite the fact that the differences between these bar-like representations merit an entirely different interpretation of centre and spread. According to Cooper and Shore (2010), it is only recently that more attention has been given to distinguishing between these graphs.

TABLE 1: The number of books read by 12 students.

Name of child	Mary	Anne	George	Barb	Tom	Jerry	Dan	Laura	Carol	Fred	Ken	Pat
No. of books read	2	4	4	4	3	0	2	3	4	2	1	1

Source: Watson, J., & Fitzallen, N. (2010). *The development of graph understanding in the mathematics curriculum: Report for the NSW Department of Education and Training*. Sydney: NSW Department of Education and Training. Available from http://www.curriculumsupport.education.nsw.gov.au/primary/mathematics/assets/pdf/dev_graph_undstdmaths.pdf

**FIGURE 2:** Number of books read by each of the 12 children.

Watson and Fitzallen (2010) use the following example to show the links and differences between these bar-like representations: 'In a class of 12 children a survey was taken to find out how many books each child read. The results of the survey then generated the ... data [shown in Table 1]'

Value bar graph

Cooper and Shore (2010) argue that the simplest and perhaps the most popular way in media and research articles would be to represent such data as shown in Figure 2.

Such a representation is often encountered by learners as early as preschool and is typical of the way in which data is represented in elementary and middle school curricula. Without discrediting other terms that have been used elsewhere, in this article I will refer to it as a value bar graph consistent with Cooper and Shore's (2010) terminology. Similarly, records of rainfall throughout the year are usually presented in such value bar graphs with the vertical axis showing the amount in centimetres or inches and the horizontal axis showing the months of the year from January right through to December as in Figure 3.

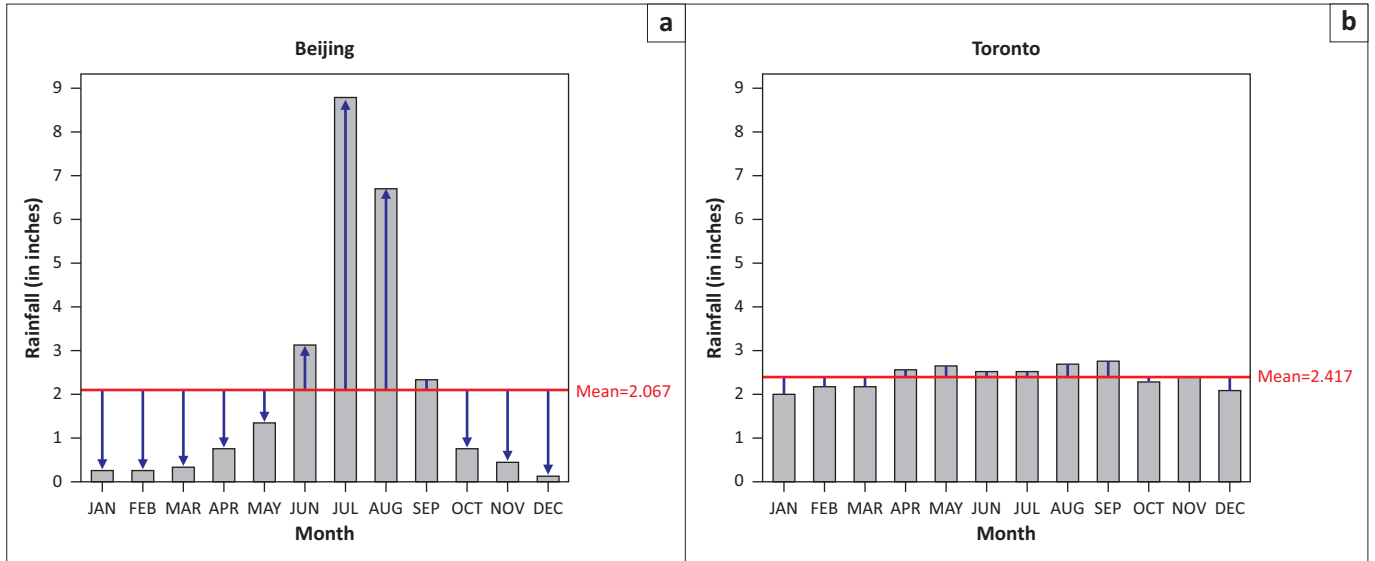
The critical distinguishing features in both cases (Figure 2 and Figure 3) are that bars represent values of single cases (number of books read by each child or the amount of rainfall that fell in each month) and in both cases the mean can be interpreted as the height at which all bars would be level as shown with the superimposed horizontal line in Figure 3. One might notice that even the most rudimentary measure of variability (the range) is also perceived on the vertical axis (difference between the highest and lowest bars). Other measures of variability in the data are also perceived through the vertical axis and would then be judged by deviations from the mean – the superimposed horizontal line in Figure 3. Notice that this superimposed horizontal could also have

been drawn in Figure 2 to enable visualisation of variability from the mean number of books read. Admittedly such a representation would only be useful when dealing with a small number of cases or data, hence such 'value bar graphs' are suitable in elementary and middle school work. Cooper and Shore (2010) warn of misconceptions that manifest when this correct perception in a value bar graph is juxtaposed onto other more complex bar-like representations, resulting in learners incorrectly interpreting such measures. In order to appreciate this difference in perceiving variability in data, let us look at how the distribution bar graph is developed from such a value bar graph.

Distribution bar graph

Let me point out here that, historically, bar-like representations are rooted in geographical analysis of population statistics where a large amount of information was gathered (Cooper & Shore, 2010). Despite the fact that different data representation techniques have been developed over the years the goal in data handling remains focused on analysis of large multivariate data sets; hence, learners should develop the skills of dealing with summaries (not cases) of large amounts of information. The same example of the number of books read by 12 children is used to show the transition from a value bar graph to a more complex distribution bar graph which aggregates data. Looking across the data in Table 1, there are five possible values the data could take: 0, 1, 2, 3 and 4. It is important to note that just like we could write the children's names in any order so we could also write the values in any order because in this context these are mere labels. The frequencies for each value are determined by the counts of children who read that number of books, as in Figure 4.

The resultant graph is an aggregation of data (distribution bar graph) as opposed to single cases that characterise a value bar graph. We immediately notice how in the distribution bar graph, the individual cases are lost as we can no longer tell how many books were read by each of the children. According to Cooper and Shore (2010), these two types of graphs (value bar graph and distribution bar graph) may superficially look the same. Both have qualitative values (categories or case names) usually on the horizontal axis and numerical scale on the vertical axis. In each case the height (or length) of the bars represents the value of the data counts. However, the difference between the two graphs is that each 'bar' for a value bar graph represents data associated with an individual (number of books read by each child) whereas a distribution bar graph collects together number of books read and reports their total frequency. They also differ in that, visually, the method to judge variability is exactly the opposite. For example, the



Source: Cooper, L.L., & Shore, F.S. (2010). The effects of data and graph type on concepts and visualisation of variability. *Journal of Statistics Education*, 18(2), 1–16

FIGURE 3: Rainfall for Beijing and Toronto

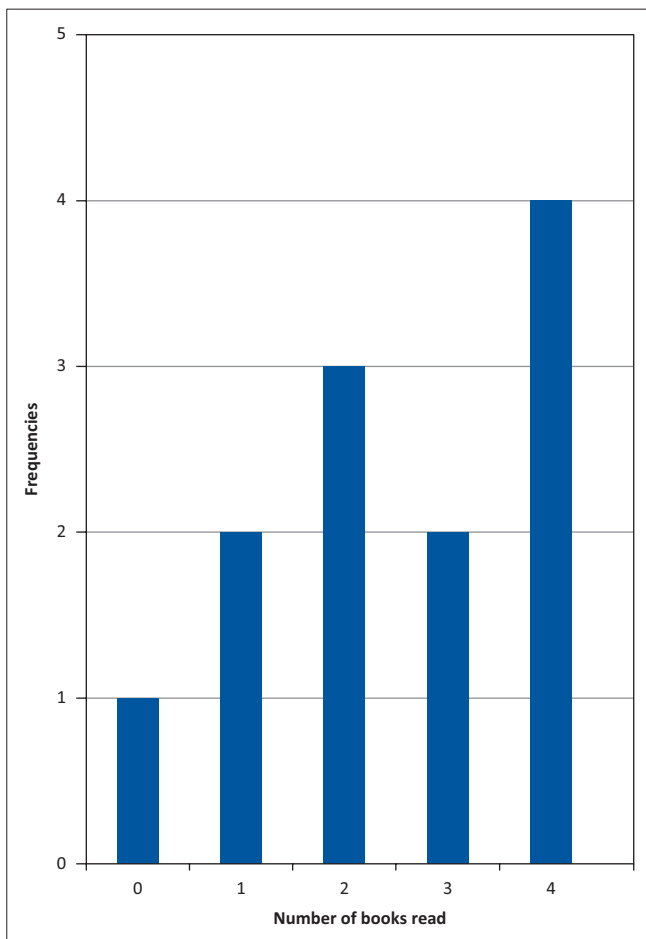


FIGURE 4: Distribution bar graph for the number of books read by 12 children.

highest bar in a value bar graph measures the maximum score (highest number of books read by a learner) whereas the highest bar in a distribution bar graph measures the mode (the number of books read by most learners). These are clearly different measures, the former being a measure of variability and the latter being a measure of centre.

To elaborate further on this point, if we superimposed a horizontal line for the mean (the height at which all bars would be level) in the value bar graph (Figure 2 and Figure 3), variability in the data (how far above and below the mean) is perceived through variation in the bar heights. On the other hand the centre for a distribution bar graph implies a typical categorical value (modal) found on the horizontal axis. Furthermore, in the case of the distribution bar graphs, bars of approximately equal height indicate great variability, whereas for value bar graphs, the same visual display of approximately equal bar heights indicates little variability. So in summary, we notice immediately that in distribution bar graphs, measures of centre and variability are no longer perceived from the vertical axis as in the case of the value bar graph. For data sets that have a typical value (mode), the greater the frequency of that modal category compared to frequencies of other categories, the more alike the data are and thus the less variable the data. The more the data differ from the modal category, to the extreme point that there is no longer a concentration of values, the more variable the data. The extent to which the modal category’s frequency stands apart from the frequencies of other categories therefore determines the appropriateness to refer to the mode as a typical value (Cooper & Shore, 2010).

The histogram

Within the group of bar-like representations, the histogram is an innovation developed from the distribution bar graph. According to Cooper and Shore (2010), its use of bars makes the histogram visually similar to the two other types of graphs (value bar graphs and distribution bar graphs) discussed earlier and thus it can potentially be confused with them. Categorical scales come in three fundamental types: nominal, ordinal and interval. Whilst value bar graphs and distribution bar graphs usually plot nominal and ordinal data respectively, in a histogram, each bar represents the frequency of intervals of continuous

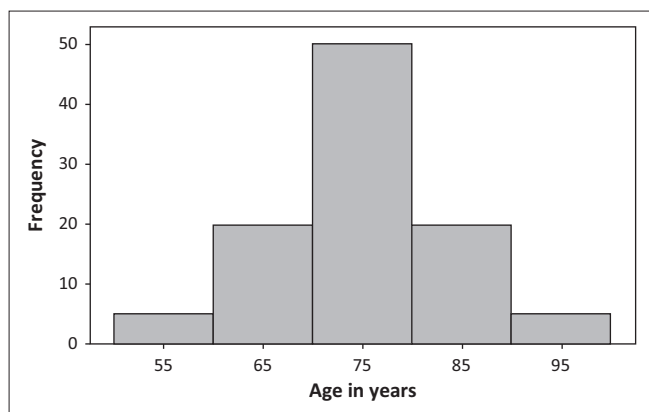


FIGURE 5: A histogram showing the distribution of ages of people in a region.

data. I will use an example to illustrate how histograms represent continuous data.

Let us say we want to count the number of people in a region who are aged 50 years and older. However, we might not want to report a separate count for every individual case of the 1000 people that fall within this age range (a value bar graph) and neither do we want to report on an individual age from 50 to 100 (a distribution bar graph). This age range (50–100) could then be converted into interval scale by subdividing the full range into smaller ranges, for example, ranges labelled 50–59, 60–69, 70–79, 80–89, and 90–99. According to Few (2005), an interval scale starts out as a quantitative scale that is then converted into a categorical scale by subdividing the range of values into a sequential series of smaller ranges of equal size (intervals) and by giving each range a categorical label. Age is a typical example of a continuous variable and in Figure 5 we see how the histogram summarises the data.

Histograms are best used with data where non-integers are actually possible; hence the bars are drawn adjacent to each other as they represent intervals of continuous data. The numbers on the horizontal axis correspond to the midpoints of the intervals (e.g. 55 in the first interval of 50–60), which determine where a particular data point gets counted on the histogram. Due to the use of the midpoint value the raw data values are no longer accessible in a histogram. The reader therefore is less likely to calculate a measure of variability and even when an attempt is made, accuracy is lost in measures of centre such as the mean as they become more estimates. In a histogram the counting of a particular data point at the midpoint of intervals is supported by Cooper and Shore (2010), who argue that at times we may want to read the trend of the distribution. We can achieve this by creating a histogram or frequency polygon from a histogram. A frequency polygon displays data by using line segments connecting points plotted for the frequencies at the midpoint of each class interval. A histogram is used only when depicting data from the continuous variables shown on a histogram. Given these conventions, the analysis then focused on the extent to which learners' representations were consistent with or in violation of these conventions.

Methodology

Participants

This article works with archived data collected from four experienced (over seven years on average) Grade 11 teachers, two male and two female (Mhlolo & Schäfer, 2012). Twenty lessons on number, algebra and data handling topics were video recorded and transcribed, generating a 300-page database. This article focuses on the lessons from one male Grade 11 teacher who was observed teaching data handling to a class of 43 learners. Prior to the lessons, the learners had been tasked by this teacher to collect data on the number of children in different households around the school. This was for a Mathematics project which formed part of their curriculum requirements. The lessons from which this article draws data could be described as learner-centred in that the teacher took more of a back seat and wanted to see how the learners would handle the data they had collected. This presented an ideal environment for the researcher to understand how the learners assimilated their prior knowledge in a typical problem-solving situation. The lessons were demarcated into four units of analysis and the criteria for demarcation are briefly discussed.

Units of analysis

There is general consensus on the view that learners' meta-representational competence is the very resource out of which expertise is constructed (diSessa & Sherin, 2000) and a number of researchers have used other terms such as phenomenological primitives (p-prims) (diSessa, 1993, 2004), cues (Davis, 1984) or 'met before's' (Tall, 2008) in support of the existence of such a pool of knowledge. Kosslyn (1989) suggests that in order to analyse learners' meta-representational competence for graphs it is necessary to examine their understanding of the interrelated connections amongst three broad constituents of a graph: a frequency distribution table, a framework and a specifier. Consistent with this suggestion, in this article, Analysis Unit 1 focuses on how learners created the table for the graph, Unit 2 on how they drew the axes and Unit 3 on construction of the bars. Unit 4 was added to focus on the final bar-like representation that was drawn by learners. Whilst connections between these interrelated constituents of a graph are necessary, an observation made by Few (2005) was that most people walk through these choices as if they were sleepwalking, with only a vague sense of what works or why one choice is better than another.

Data

We pick up the conversation after the learners had drawn a frequency table on the board showing the results of the survey of the number of children in different households. Initially the table had been drawn without the tally column. In the extracts below, 'T' stands for teacher, 'L' for learner and 'Chorus' indicates a group response.

Unit 1: Construction of a table

- T: So what do we do next after you have drawn the frequency table?
- Chorus: We make tallies. We make a pie chart. We make a graph. *[After a while it is agreed that the table should have tallies.]*
- L1: *[Comes to the board and makes a tally of the number 8 as requested by the teacher.]*
- T: Have you ever seen something like this?
- Chorus: Yes
- T: Where?
- Chorus: Last year. Last of last year. The previous maths teacher.
- T: So the previous maths teacher showed you how to tally? OK, can you complete the table then. *[The table is then completed as shown in Figure 6.]*

Unit 2: Drawing the axes

- T: Now after this information, how can you display this information? What it is like here, the information has been collected and now it has been organised. OK now how are you going to display this information?
- L2: In a graph.
- T: Graph, we have different types of graphs and also we have different types of data. It's grouped and ungrouped. The way you display grouped data is not the same way as you display ungrouped data. So what type of a graph?
- L3: Bar graph.
- T: Can somebody show us how to go about it?
- L4: *[Comes to the board and draws two axes labelled as in Figure 7.]*
- T: OK what do you call this line? *[Points to the horizontal axis.]*
- L5: The x-axis.
- T: Now on the horizontal or the vertical OK you need to have either the number of children in each family and on the other you need to have maybe type of frequency whatever.
- L6: *[Comes to the board and labels the horizontal axis as 'number of children in different families'. The vertical axis is labelled as the frequency axis.]*

No. of children	Frequency	Tally
0	8	
1	14	
2	20	
3	17	
4	10	
5	11	

FIGURE 6: Frequency table for the number of children in different households.

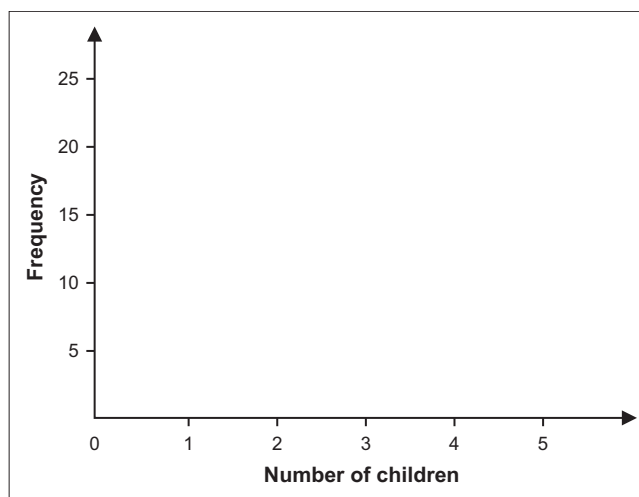


FIGURE 7: Axes drawn for the bar graph.

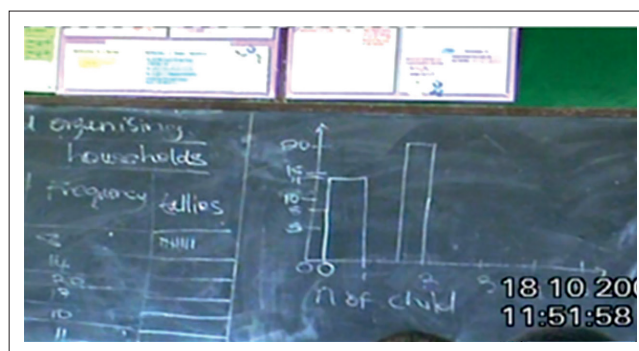


FIGURE 8: Graph showing the number of children in different households.

Unit 3: Construction of the bars

- T: Now how are you going to display your data? Where, OK here it is number of children *[pointing to the horizontal axis]*. We start with what? Now because it's a bar graph, how would you put it here? Like this is the bar *[teacher drawing examples of horizontal and vertical rectangular blocks]*. Now how are you going to display your 0 and 8?
- L7: *[Comes to the board and draws the first bar in between 0 and 1 on the horizontal axis in Figure 8.]*
- T: Is he correct?
- Chorus: Somehow, almost, maybe.
- L8: That bar shows a quarter and eight, Ma'am.
- T: OK so the zero was supposed to be where? Here?
- L9: *[Goes to the board and places a second 0 at the point where the bar intersected with the horizontal axis making the first bar sit between the two zeros as shown in Figure 7 and Figure 8.]*
- T: So if the 0 was here he would be correct.
- L10: Maybe it's incorrect.
- L11: It's incorrect.
- T: Let's see if you can put the bar for 1 and 14. Let's see. Let's try.
- L12: *[Comes to the board and places the second bar showing a frequency of 14 as shown in Figure 8 and Figure 9.]*

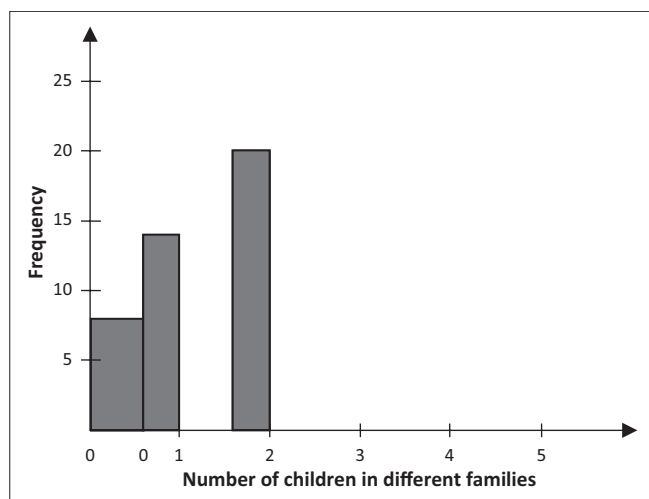


FIGURE 9: Learners' bar-like graph of survey data.

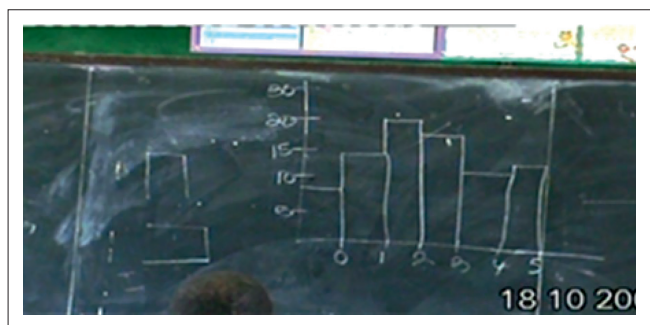


FIGURE 10: Bar-like graph of the number of children in different households.

L13: [Commenting after the second bar had been drawn] It's wrong.

T: [Asks yet another learner to draw the bar showing 2 and 20. An interesting observation made is that whilst the first two bars had been drawn adjacent to each other, this third bar was disjointed as shown in Figure 8. The graph was re-drawn for clarity (Figure 9).]

[After some long discussions on whether or not the graph was representing the data accurately, it was erased.]

L14: [Comes to the board and draws another new set of axes. The zero which was at the intersection of the horizontal and vertical axes is then removed leaving the second zero and the other values as they were on the abandoned axes.]

Chorus: [Learners take turns to draw the bars on this new set of axes as shown in Figure 10.]

Unit 4: The final bar-like representation

T: If you were to display something like this to a person who doesn't know mathematics will that person be in a position to read? OK remember that you have organised your data and now you are displaying your data, can a person be able to read this?

L15: I think maybe you have to label whether which side is talking about number of children and the households. [This comment came because the axes had not been labelled.]

T: Ok now turn to the notice board. Look at the graph of inflation. This type of graph is called a bar graph. Look at it and the one we have just drawn. What is the difference? [There was a graph in class showing inflation rates from 1999 to 2009.]

Chorus: The spaces, it's decorated, it's neatly displayed. [Lesson ends]

Discussion

The questions steering this analysis were:

1. What is the nature of learners' prior knowledge for graphs?
2. Where do these ideas come from?
3. How are they involved both productively and unproductively in the process of constructing bar graphs?

Each unit of analysis attempts to answer these three questions.

Unit 1: Constructing the frequency table

From the discussion that took place during the process of making a frequency table for the collected data, it is evident the learners brought the knowledge of tallying from the 'previous teacher'. It can be argued that the knowledge of tallying was neither supportive nor problematic since with or without the tally column the students would still have been able to construct a correct bar graph. However, the agreement by learners that frequencies should be 'tallied' opened up a number of questions about their procedural and conceptual understanding of tallying. Let us recall that a tally is a mark used in recording a number of acts or objects, most often consisting of four vertical lines cancelled diagonally or horizontally by a fifth line. Tallying or counting is the act of finding the number of elements of a finite set of objects through a one-to-one correspondence. It is meant to avoid visiting the same element more than once. After tallying the value of the final object gives the desired number of elements (cardinality) in that set. So if the learners' frequency table had a column of frequencies, by implication tallying had already been done. Therefore from the learners' wanting to tally the number 8 or 14 or 20 (frequencies) it can be concluded that the purpose of tallying and when it should be done were not clear to them. This suggests that learners had a superficial understanding of the concept.

Unit 2: Drawing the axes

When prompted to show the information on a bar graph, what is evident is that learners brought their prior knowledge of a framework of a graph with an x -axis and a y -axis intersecting at 0 and scaled on both axes as shown in Figure 7. Students meet this type of framework more often when solving equations graphically. Was this prior knowledge supportive? To a certain extent this prior knowledge was supportive for, according to Friel *et al.* (2001), graphs share similar structural components. The framework of a graph as discussed earlier gives information about the kinds of measurements being



used and the data being measured. The simplest framework has this L-shape that learners drew, with one leg (x -axis) standing for the data being measured and the other leg (y -axis) providing information about the measurements that are being used. This was important for the learners to be able to represent their data on a bar graph.

However, to a larger extent, it is evident that their prior knowledge of axes was not very productive as they later struggled to draw the bars for their data. When both the x -axis and the y -axis have numerical information, as was the case in this task, learners needed to have a deeper knowledge of numbers in order to figure out which numerical information goes onto which axis. Curcio (1987) reports that the mathematical contents of a graph, that is, number concept, relationships and fundamental operations contained in it, were factors in which prior knowledge seemed necessary for graph comprehension. The recommendation was that the relationship between the subject matter of number and choice of graph form should be further investigated.

It is evident that learners did not have a clear understanding of this relationship. By drawing a framework of a graph with an x -axis and a y -axis intersecting at 0 and scaled on both axes learners implied a functional relationship between the variables depicted on the axes. Yet bar graphs by convention are not used to convey functional relationships (Follettie, 1980) because such a graph of categorical data displays the relative magnitudes without implying a functional relationship. Therefore, conventionally a bar graph of categorical data would have a scale only on its frequency axis. In a similar study on high school and college students, delMas, Garfield, Ooms and Chance (2007) also speculate that learners do not actually understand what the axes represent. Friel and Bright (1995) caution that interpreting graphs that utilise two axes may present difficulties if the nature of data that they represent across different graphs is not explicitly recognised. When considering graphs with any of these frameworks as tools for data reduction, one should note the differences in the nature of data that are represented on these axes. In the case of a value bar graph, distribution bar graph or histograms, the major difference is in what is represented on the x -axis. For example, in a value bar graph drawn with vertical columns, the columns are positioned over a label on the x -axis that represents a nominal measure. A nominal measure refers to data that consist of names or categories so that the data cannot be arranged in any specific ordering scheme. The nominal level of measurement occurs when the observations do not have a meaningful numeric value, for example numbers assigned to soccer players. The values of nominal variables cannot be meaningfully compared to see if one is larger than another, cannot be added, subtracted, multiplied or divided nor can the mean be calculated (what most people call the average). So in this case, the x -axis does not have a low end or a high end, because the labels on the x -axis are categorical and not quantitative. Learners get experience of such categorical

bar graphs much earlier than functional graphs. They draw graphs of weather in a week where the horizontal axis is labelled with the days of the week as early as Grade 1. So one can argue that learners' pre-knowledge of symbolic functional graphs where the numbers on the x -axis represent a scale like on a number line was a stumbling block to understand how to represent categorical data as labels without a scale or order.

Unit 3: Constructing the bars

After drawing the axes, it was evident that learners did bring their prior knowledge of matching the height of bars with the frequencies (see Figure 8 and Figure 10). Generally a bar graph plots the number of times a particular value or category occurs in a data set, with the height of the bar representing the number of observations of that score or that category. It is evident from Figure 8 and Figure 10 (see marks placed between 5 and 10 and 10 and 15 on the vertical axis) that this knowledge was productive in terms of matching precisely the height of bars for 8, 14 and 20 with the frequencies.

The problem however surfaced in terms of where these bars sit. By placing the 0 at the origin the class struggled to draw the first bar showing 8 families with 0 children each and the subsequent bars were also problematic. This suggests that learners were unable to distinguish the data set that they were dealing with. Distinguishing between sets of data as discrete cases, discrete categories or grouped numerical data along some scale is a critical factor for constructing appropriate representations of the data. In all the three representations of categorical data, that is, value bar graphs, distribution bar graphs and histograms, categories of the variable are typically marked at the midpoints of the category on that particular axis (horizontal if it is a column graph and vertical if it is a bar graph). From the way learners drew their bars, it is evident that this convention was not recognised as their bars were sitting on two different numbers at the same time.

Another evident failure to recognise a convention was that at times learners drew joint bars as in the histogram and at times disjoint bars as in a bar graph, yet conventionally histograms must have joint bars and bar graphs must have disjoint bars. A study on learners' conceptual understanding of statistics by delMas *et al.* (2007) identified learners' inability to recognise critical differences between histograms and other graph types that use bars. This would have been expected given that empirical evidence shows that at school level these graphs are usually referred to as bar graphs and only recently has more attention been given to distinguishing between these graphs (Cooper & Shore, 2010).

Unit 4: The final representation

Let us recall that the learners wanted to represent their own collected data on a bar graph. The question then is: to what extent did they achieve this objective? We notice



from the basic-level constituents discussed earlier that the learners' representation is neither a value bar graph, nor a distribution bar graph nor a histogram. Whilst the heights of bars matched with the frequencies, they were joint bars and were sitting on two different values on the horizontal axis in violation of the midpoint convention that guides where bars should be located in value bar graphs, distribution bar graphs and histograms. The overall mathematical outcome here was something close to a histogram but did not represent the original data set particularly well, either in terms of mathematical structure and convention or with reference to the real-world situation being represented. This suggests that learners' meta-representational competences were inadequate for bar graph construction.

When numbers are used in bar graphs, the axis that assumes a categorical scale could represent three fundamental types: nominal, ordinal and interval data. These categorical contexts of number are problematic even with adults given that the majority of time spent on number and operations in the earlier grades focuses on numbers in their quantitative contexts, with learners usually encountering the categorical contexts of number only when dealing with data handling tasks. This suggests that to communicate effectively using graphs, one has to understand the nature of the data, graphing conventions and a bit about visual perception. Without guiding principles rooted in a clear understanding of graph design, choices are arbitrary and the resulting communication fails to represent the information effectively, as was the case in this class.

Implications

This article has both theoretical and practical implications. In terms of theory, this article has shown that due to the more recent emergence of the field of statistics, there is much more flexibility in nomenclature and lack of convergence on what the conventions should be. Watson and Fitzallen (2010) show how for example at both primary and high school levels, these bar-like representations are often simply referred to as bar graphs so that their distinction is unclear. Yet from this article it has been shown that the methods of judging both centre and variability are clearly different across such bar-like representations. Cooper and Shore (2010) show how an understanding of measures of centre and variability was the single most important foundational concept in all statistical thinking. So in order to teach these concepts effectively, curricula need to be constructed and implemented carefully; writing realistic assessment items plus having the resources to mark them is not easy if graphs continue to be referred to loosely as bar graphs. All this points to the need to converge on some specific naming of these bar-like representations and this article suggests that Cooper and Shore's way of distinguishing between value bar graphs, distribution bar graphs and histograms guides us towards such convergence in nomenclature.

In terms of concept formation, as long as these bar-like representations are referred to loosely as bar graphs, learners

will not make connections between the different graphical representations of quantitative data and their corresponding ways of conveying information on measures of centre and variability for that data. Research indicates that learners entering college may have only a superficial understanding of centre and variability and are likely to have particular difficulty extracting information about those measures when data are presented in graphical form (Cooper & Shore, 2010). Yet Franklin *et al.* (2007) maintain that an understanding of variability in data is the single most important foundational concept in all of statistical thinking. A solution to this problem might be addressed by this convergence in conventions as suggested in this article.

In terms of practice, this study argues that knowing the ways in which these types of bar-like graphs (value bar graphs, distribution bar graphs and histograms) represent certain types of data may help teachers make decisions about the level of complexity for instruction. Whilst the so called 'bar graph' is often encountered by students as early as preschool, this article argues that the level of complexity of categorical data that is handled by learners at that early stage is low. This is the kind of data that is best represented in what has been defined in this article as the value bar graph. Friel *et al.* (2001) show that the transition from these case value bar graphs to distribution bar graphs may be confusing if this transition is not carefully considered and explored because the axes must be redefined. This confusion is evident in this article: learners wanted to draw a bar graph but they ended up with something close to a histogram, suggesting that they could not distinguish between these types of bar-like graphs. The view is that teachers should create a gradual transition from drawing graphs with objects themselves (value bar graphs) to the more abstract distribution bar graph (Rangecroft, 1994). A similar suggestion put forth by Franklin *et al.* (2007) was that both primary and secondary learners engage in tasks that require them to integrate deep understanding of graphical representation along with measures of centre and spread through a steady progression from value bar graphs, through distribution bar graphs to histograms.

Acknowledgements

I acknowledge the Department for International Development for funding the PhD study from which this article is drawn. The views expressed in this article are not necessarily those of the funders.

Competing interests

The author declares that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Ethical considerations

The Department of Education granted approval to proceed with this study under permit T-728 P01/02 U-848.



References

- Cooper, L.L., & Shore, F.S. (2010). The effects of data and graph type on concepts and visualisation of variability. *Journal of Statistics Education*, 18(2), 1–16.
- Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382–393. <http://dx.doi.org/10.2307/749086>
- Davis, R.B. (1984). *Learning mathematics. The cognitive science approach to mathematics education*. London: Croom Helm.
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.
- diSessa, A.A. (1993). Towards an epistemology of Physics. *Cognition and Instruction*, 10(2–3), 105–125. <http://dx.doi.org/10.1080/07370008.1985.9649008>
- diSessa, A.A., (2004). Meta-representation: Native competence and targets for instruction. *Cognition and Instruction*, 22(3), 293–331. http://dx.doi.org/10.1207/s1532690xci2203_2
- diSessa, A.A., Hammer, D., Sherin, B.L., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior*, 10, 117–160.
- diSessa, A.A., & Sherin, B.L. (2000). Meta-representation: An introduction. *Journal of Mathematical Behavior*, 19, 385–398. [http://dx.doi.org/10.1016/S0732-3123\(01\)00051-7](http://dx.doi.org/10.1016/S0732-3123(01)00051-7)
- English, L. (2012). Young children's meta-representational competence in data modelling. In J. Dindyal, L.P. Cheng, & S.F. Ng (Eds.), *Mathematics education: Exploring horizons (Proceedings of the 35th Annual Conference of the Mathematics Education research Group of Australasia)* (pp. 266–273). Singapore: MERGA. Available from http://www.merga.net.au/publications/counter.php?pub=pub_conf&id=1959
- Few, S. (2005). Quantitative vs. categorical data: A difference worth knowing. *Perceptual Edge*, April, 1–5.
- Follettie, J.F. (1980). *Bar graph-using operations and response time (Technical Report)*. ERIC Document Reproduction Service No. ED 250 381. Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Alexandria, VA: American Statistical Association.
- Friel, S., & Bright, G. (1995). Graph knowledge: Understanding how students interpret data using graphs. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.
- Friel, S.N., Curcio, F.R., & Bright, G.W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158. <http://dx.doi.org/10.2307/749671>
- Kosslyn, S.M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–226. <http://dx.doi.org/10.1002/acp.2350030302>
- Mhlolo, M.K., & Schäfer, M. (2012). Towards empowering learners in a democratic mathematics classroom: To what extent are teachers' listening orientations conducive to and respectful of learners' thinking? *Pythagoras*, 33(2), 79–87. <http://dx.doi.org/10.4102/pythagoras.v33i2.166>
- Rangecroft, M. (1994). Graph work – Developing a progression. In D. Green (Ed.), *The best of teaching statistics* (pp. 7–12). Sheffield: The Teaching Statistics Trust.
- Shah, P., & Hoeffner, J., (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47–69. <http://dx.doi.org/10.1023/A:1013180410169>
- Tall, D. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal*, 20(2), 5–24. <http://dx.doi.org/10.1007/BF03217474>
- Watson, J., & Fitzallen, N. (2010). *The development of graph understanding in the mathematics curriculum: Report for the NSW Department of Education and Training*. Sydney: NSW Department of Education and Training. Available from http://www.curriculumsupport.education.nsw.gov.au/primary/mathematics/assets/pdf/dev_graph_undstdmaths.pdf