SHOULD YOUNG STUDENTS LEARN ABOUT BOX PLOTS?

Arthur Bakker

Freudenthal Institute,

Utrecht University, the Netherlands

arthur@fi.uu.nl

Rolf Biehler

University of Kassel,

Germany

biehler@mathematik.uni-kassel.de

Cliff Konold

SRRI, University of Massachusetts,

Amherst, MA, USA

konold@srri.umass.edu

Paper for Roundtable, June 27 – July 3, 2004, in Lund, Sweden; presented by the first author. The authors are listed in alphabetical order.

Version of April 29, 2004

SHOULD YOUNG STUDENTS LEARN ABOUT BOX PLOTS?

Arthur Bakker
Utrecht University
The Netherlands

Rolf Biehler
University of Kassel
Germany

Cliff Konold
University of Massachusetts, Amherst, USA

Abstract

In this paper, we explore the challenges of learning about box plots and question the rationale for introducing box plots to middle school students (up to 14 years old). Box plots are very valuable tools for data analysis and for those who know how to interpret them. Research has shown, however, that some of their features make them particularly difficult for young students to use in authentic contexts. These include that

a) box plots generally do not allow perceiving individual cases;

b) box plots operate differently than other displays students encounter;

c) the median is not as intuitive to students as we once suspected;

d) quartiles divide the data into groups in ways that few students (or even teachers) really understand.

We recommend that educators consider these features as they determine whether, how, and when to introduce box plot to students.

Introduction

Figure 1 shows box plots of a data set from a questionnaire that 540 students of grade 11 responded to (Biehler, 2003). The students were asked when they usually go to bed and when they get up in the morning. From these data, total sleeping time (in hours) was calculated. “Sunday” refers to the hours sleeping from Saturday night to Sunday morning, and so on. The total times sleeping on weekdays are remarkably similar to one another, whereas the weekend times appear strikingly different. Students appear to sleep longer on weekend nights (although they go to bed later). Furthermore, the spread in weekend nights is larger than during the weekdays, pointing to a larger diversity of students’ habits on weekend nights where they are free of the constraints of weekday nights such as school start time and parental regulation.

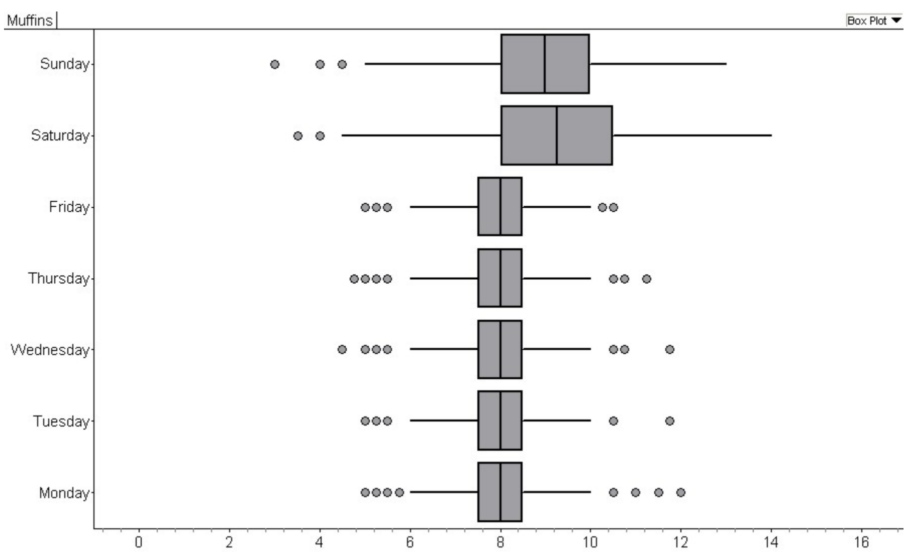

Figure 1. Box plots sleeping hours of eleventh graders for the different nights of the week (created in Fathom).

Box plots are a powerful display for comparing distributions. They provide a compact view of where the data are centered and how they are distributed over the range of the variable. They also provide easy ways to compare parts of distributions to see, for example, how the data in the top quartile compares in two groups. In several countries, box plots have become part of the standard data analysis curriculum, but the age of students to whom they are introduced differs considerably. In the USA, students learn about box plots as young as about age 12. The National Council of Teachers of Mathematics, for example, includes box plots among the list of displays that students in grades 6-8 should become familiar with (NCTM, 2000). In other countries that we know of, box plots are introduced at a somewhat later age or not at all.

Despite all the advantages in using box plots to analyze data, we think there are several features of box plots that pose particular challenges to students. These features include that:

a) box plots generally do not allow perceiving individual cases;

b) box plots operate differently than every other display students encounter;

c) the median is not as intuitive to students as we once suspected;

d) percentiles divide the data into groups in a way that few students (or even teachers) really understand.

In this paper, we first briefly review the origin of box plots and their place in the curriculum of various countries. Then we elaborate each of the features listed above and cite, when we can, what we have learned from research with students. Unfortunately, there has not been much research on box plot interpretation to date. We therefore include experiences from teaching experiments that we have carried out ourselves. We question the wisdom of introducing students as young as age 12 to box plots and recommend that educators consider the features we describe as they determine whether, how, and when to introduce box plot to students.

Origin of box plots: exploratory data analysis

Box plots are part of a general tool kit of techniques of Exploratory Data Analysis (EDA), a relatively new field of statistics in which data are explored with graphical techniques (Tukey, 1977). Unlike traditional inferential statistics, the goal in EDA is not to test specific and preformed hypotheses with data from randomly drawn samples. Rather, EDA focuses on the detection of unanticipated patterns and trends in data of all types, whether randomly sampled or not. At the time Tukey introduced EDA, even statisticians rarely looked at graphs of their data, because they were so time consuming to construct. Tukey (1977) developed various paper-and-pencil methods of graph construction to encourage the use of graphic displays for the purpose of analyzing data. Famous in this regard are the stem-and-leaf plot and the box plot. Even though today's computer software can handle the labor of graph constructions, box plots, for example, are still used by experts as a powerful way to visually summarize the center and spread of distributions.

EDA has been widely adopted by statistics educators in large part because it serves the need for more data and what we can learn from them, and does not focus on the underlying theory and complicated recipes (Biehler & Steinbring, 1991; Cobb & Moore, 1997; Scheaffer, 2000). There are probably several reasons why educators in the USA decided to introduce students as young as age 12 to box plots. First, the box plot incorporates the median as the measure of center, and some early research had suggested that the median is easier for students to understand as a measure of center than is the mean (Mokros & Russell, 1995). Box plots also provide in the Interquartile Range (IQR) a measure of the degree of spread, an alternative to the computationally more challenging standard deviation. (Besides, a clear geometrical interpretation of the SD can only be developed in the context of normal distributions.) Furthermore, box plots depict both the measure of spread and center pictorially, which is largely why box plots are such a powerful way to quickly compare several groups at once. Therefore the box plot and the interquartile range promised to provide better tools for developing an initial feeling for spread than other graphs and measures of spread.

Box plots in the curricula of various countries

There are large differences in the mathematics and statistics curricula of various countries. As we mentioned, box plots are introduced to 12-year-old students in the USA, but in other countries, they are introduced somewhat later: in New Zealand to 13-14-year-olds, in Australia, Belgium, the Netherlands, South Africa to 15-16-year-olds, and in France to 16-17-year-olds. In Israel, box plots are not in the secondary school curriculum.

A group of stochastics educators in Germany (Arbeitskreis Stochastik, 2003) recommend the use of box plots for students of 15 years old and above. Currently, however, more than half of the German states introduce neither box plots nor the median to secondary students. In fact, most states introduce no descriptive statistics before students are 15.

In the next four sections we elaborate on the features A to D mentioned earlier, which form the challenges of learning about box plots.

A. Individual cases versus aggregate information

Statistics is concerned with patterns and trends that become evident in collections of cases. A number of researchers have found, however, that students new to the study of statistics are prone to attend to individual cases, or to frequencies of cases with the same or similar value (Bakker & Gravemeijer, in press; Ben-Zvi & Arcavi, 2001; Biehler & Steinbring, 1991; Cobb et al., 2003; Hancock et al., 1992; Konold et al., 1997). We believe that one of the core challenges of statistics education is to support students in making the transition from a case-oriented view to an aggregate view of data.

If we look at graphical representations of data, we notice that with some of them, individual cases are recognizable (e.g., dot plots and scatterplots) while with others they are not (box plots and histograms). One could argue that introducing students to plots such as box plots in which only aggregate features are depicted might be an effective way to wean students from attending to individual cases. In our experience, however, the more frequent result is to add to students' confusion. For example, Konold et al. (1997), and Biehler (1997) describe how students they interviewed tried to interpret box plots and histograms to help them answer a question they were exploring. Despite the fact that these students had just completed a yearlong statistics course, which included instruction in these displays, these students struggled to interpret these two representations. Some of them even attempted to identify individual cases within histograms and box plots, perhaps in an attempt to recall how the plots encoded data values. We recommend that early instruction in statistics focus primarily, if not exclusively, on plots in which individual cases are visible. When aggregate plots are introduced, we recommend that they initially be accompanied with representations that still allow students to see individual cases, as in Figure 2 where box plots are overlaid on top of stacked dot plots, with the option of then hiding the cases displayed in the dot plot. This option is available in recently developed educational software such as the Minitools (Cobb et al., 1997) and Tinkerplots (Konold & Miller, 2004). The Minitools form a series of three applets specially designed for middle school students. The software Fathom has the more general option to represent statistical measures as vertical lines in a dot plot.

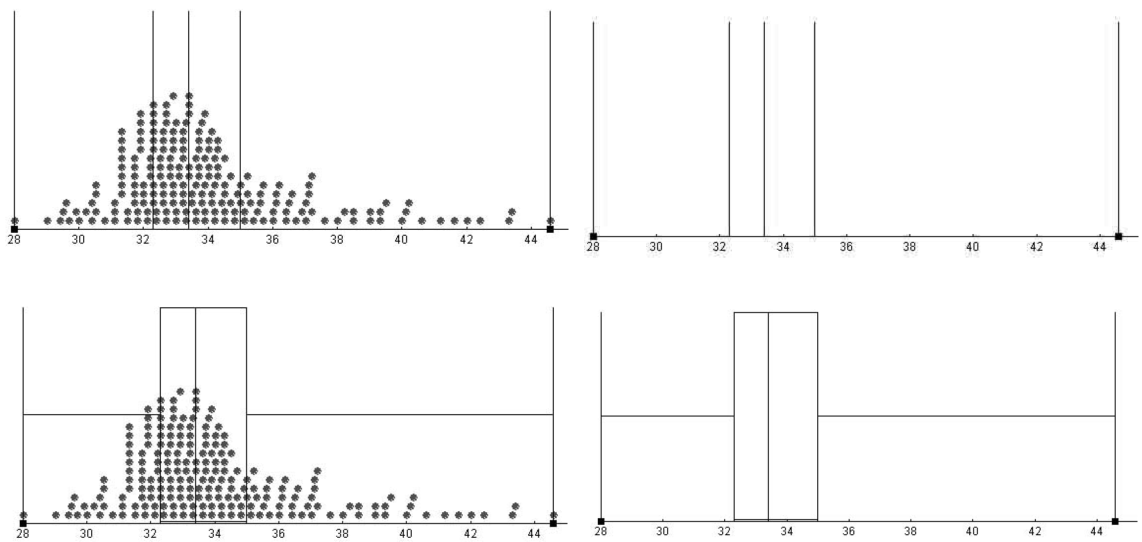

Figure 2. Option in Minitool 2 to make four equal groups (top row) 2, to hide data (right-hand graphs), and add a box plot overlay (bottom row) 2

Figure 2 is taken from Minitool 2, which offers different ways of grouping data sets such as into four (roughly) equal groups. This allows students to compare data sets with the same range and center but different spread. Note too that this option is a precursor to the box plot. Bakker and Gravemeijer (in press) added the box plot overlay in Minitool 2 to provide a stepwise support from unorganized data to conventional plots such as box plots. Similarly, the grouping option of fixed interval width is a precursor to the histogram. With such representations, students may come to see the shape of the data in relation to the quartiles. Similar options are available in Tinkerplots. Dot plots, which display all the individual cases, appear to be much easier for students to understand than aggregate plots such as histograms and box plots, in which data are aggregated (Konold & Higgins, 2003).

B. Displaying densities rather than frequencies

In representations such as frequency histograms and bar graphs, the area of a bar corresponds to the frequency of cases of a particular class or category. If bar A is twice the size of the adjacent bar B, we know that there are twice as many cases represented by bar A as there are by B. The same is obviously true of representations that show individual cases such as scatterplots, stem-and-leaf plots, and stacked dot plots: plot elements accumulate in direct proportion to frequency. If you want to see where the cases are most densely clustered in histograms, bar graphs, or stacked dot plots, you look for the tallest bar or stack. In scatterplots and dot plots, you can see both frequency and density directly.

This relation between plot area and frequency does not hold with box plots where each of the four major components contains roughly 25% of the data. Thus frequency is not encoded at all in box plots. By dividing the date equally among four parts, whose sizes then vary, box plot allow you to quickly see differences in density. But somewhat counterintuitively, density is inversely related to the size of box-plot components: the smaller a component is relative to the others, the more densely values are packed in that range. Thus, the portions of a distribution that are most pronounced in other graphs — e.g., the area with the tallest bars or highest density of values — are least pronounced in a box plot, where the smallest sections have the highest densities. We assume that this difference between box plots and most other displays contributes to making box plots particularly difficult for students to understand.

This does not mean that students cannot be quickly taught how to construct and read off salient values from box plots ("here is the median; the IQR is y; the range goes from x to y"). But questions that require more understanding, such as interpreting the meaning of the differences in center and spread of Figure 1 between weekdays and weekends, demand more than these rudimentary decoding skills. As we discuss below, they require that we interpret the median and IQR as group features.

C. The median as a measure of center

Box plots are conceptually rich. To understand them, interpreters need at least to know what minimum, first quartile, median, third quartile, and maximum are. Furthermore, they need to understand that the median is used as a measure of center (of a distribution); that the length of the box (not its width) is a measure of the spread of the data; and that the range is another measure of spread. To complicate matters further, there are many variants of box plots (McGill, Tukey, & Larsen, 1978); for instance as in Figure 1, the whiskers are not drawn to the extreme values.[i]

The median is not hard to find as the middle-most value of an ordered row of data values, or the mean of the middle-most two values. At an early age students can learn to count inwards from one end of an ordered sequence (e.g., students arranged in a line by height) to find the median. The median is also readily available in various educational software tools: in Minitool 2, students can divide the data set into two equal groups and in

Tinkerplots they can click on a median value button, represented by an inverted T (see Figure 3).

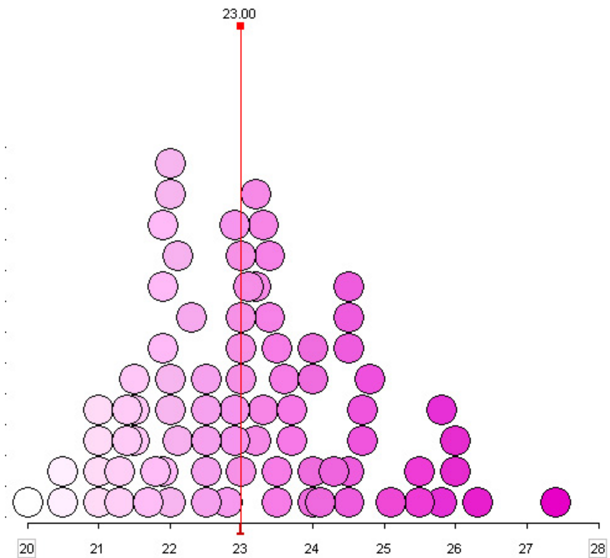

Figure 3. The median value in Tinkerplots with a 'vertical reference line' at the median (data set is about foot length in cm of sixth graders).

However, as Konold and Higgins (2003) suggest, even if students can compute the median or mean, this does not imply that they interpret these as group descriptors or measures of center. Many students tend rather to see a median or mean as a feature of an individual in the center of the group rather than as a characterization of the whole group. Cobb, McClain, and Gravemeijer (2003) observed that the eighth graders in their teaching experiment, who had considerable experience with data analysis the three Minitools (37 class periods in grade 7 and 41 in grade 8), considered the median only as a cut point in the data and not as a measure of center. A similar observation was made in the Dutch and German teaching experiments, which we describe later. As a start of statistical learning, it may be sufficient that students consider the median purely as a cut point. However, carefully designed additional activities are necessary to foster understanding the median as a measure of center and thus as a characteristic of the group.

D. Difficulties of quartiles

Quartiles are particularly tricky. Among the complexities is the issue of how to deal with the occurrence of ties (equal data values). There are different ways of doing this, and thus different definitions of quartiles. Computer programs use different definitions, and these definitions are not always well documented (Freund & Perles, 1987). We discuss a few definitions to show the complexity of quartiles and percentiles.

In an early teaching experiment of Biehler and Steinbring (1991) the 25th and 75th percentiles (quartiles) were introduced, respectively, as the median of the lower and the upper half of the data. The advantage of this definition is that once students knew how to locate the median, they could recursively apply the same technique to get the quartiles. However, a problem arises in this procedure when the number of data values is odd. How do you deal with the case that is located at the median when counting cases to locate the quartiles? Tukey (1977) included the case at the median in counting both halves, which could be one of the reason why he called them “hinges” — to distinguish them from what were usually called quartiles (see Hoaglin, 1983). So if there were five cases, all with different values, the value of the third ordered case would be the median; the values of the second and fourth cases would be the quartiles. A few teachers of the early teaching experiments (Biehler & Steinbring, 1991) reported to the researchers that their students found this double counting of the median counterintuitive. They therefore left the median out of both upper and lower groups when locating the quartiles, and thus used a different definition than their colleagues.

Software tools typically use a different definition, which is computationally easier for the computer.[ii] Regardless of the particular definition used, percentiles divide cases into groups in a way that is fairly non-intuitive. Based on our work with students and teachers, our sense is that most students, along with their teachers, believe that the lower whisker of the box plot captures exactly a quarter of the cases. Real data sets can rarely be partitioned into exactly four equal-sized groupings, whatever definition or software used.

As an illustration of the subtleties mentioned above we give an example. In Figure 4, we have formed groups in a set of data using dividers as provided in Tinkerplots. In computing the percentages that are displayed, cases that are exactly on a divider are counted in the group to the right of the divider. Using these dividers, none of the percentages on the top group could be set to 25% or 50%.

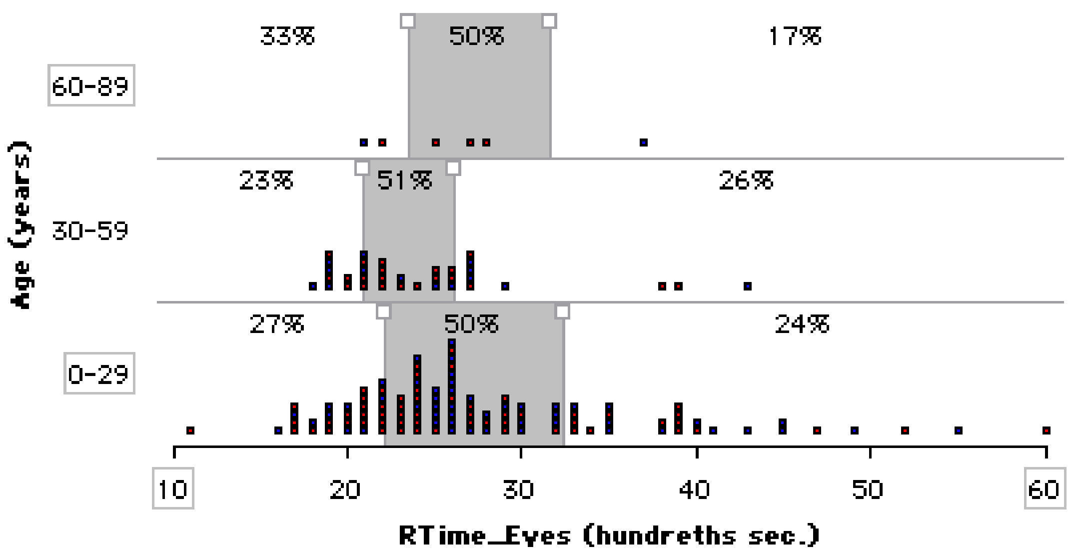

Figure 4. Reaction time (in hundreds of seconds) of different ages in response to a visual stimulus. A dot plot in Tinkerplots with the 'dividers' set as close as possible to the values one would expect in a box plot.

Quartiles do not match well the way students tend to conceive of distributions. In several recent studies, researchers noted that students tend to think of a distribution as comprising three parts, rather than four. They think about 1) the majority in the middle (which usually includes more than 50% of the cases), 2) low values, and 3) high values (Bakker & Gravemeijer, in press; Konold et al., 2002). Students also referred to the center majorities as “clumps”, which was why Konold and colleagues propose calling them “modal clumps”.

To elaborate on the issues raised in the previous sections, we briefly describe our own recent experiences drawn from teaching experiments on data analysis carried out in the Netherlands and Germany.

Teaching experiments in grade 7 in the Netherlands

Bakker (2004) carried out several teaching experiments using the Minitools in grade 7 in the Netherlands. The research built on the experiences of Cobb, McClain, and Gravemeijer (2003) in the United States. We focus on a few results concerning the median and spread.

Median. The seventh graders had several difficulties with medians. First, even though students had learned the median as the value that makes two equal groups in Minitool 2, many mistook the median for the midrange, the mean of the minimum and maximum value. Second, even if students knew that the median was the value where the data set was split into “two equal groups”, they often could not find the median in a row of numbers. This is not surprising because they had learned the median in a different graphical system, Minitool 2, with the two-equal-groups option and not as the middle-most value (or mean of the two middle-most values) in a row of numbers. Third, those students who knew the difference between mean and median sometimes interpreted them as very different from each other even when the distribution was symmetric. They did not appear to think of them as two alternative measures of a center.

Spread. The most rudimentary measure of spread, and the earliest to be used historically (David, 1998), is the range. Indeed, the seventh graders spontaneously used the range as an indication of spread, and it was often the first thing they mentioned. Next, they tended to characterize parts of data sets as “crowded, busy, full,” or “empty”, particularly when they had worked in Minitool 2 with the four-equal-groups option. They typically described a stacked dot plot in terms such as, “Here the dots are close together, and there they are spread out.” Bakker characterizes such views as ‘local views’ on spread, which we could see as precursors to density views. No student in Bakker’s study viewed spread as dispersion from a mean or median value.

One of the conclusions was that developing notions of center and spread that can be measured, respectively, with median and IQR took much more time with students of this age than was available in this teaching experiment (up to 15 lessons). Box plots could only be introduced superficially.

Teaching experiment in grade 8 in Germany

Biehler and Kombrink (1999) carried out a teaching experiment in a German comprehensive school in one class with 25 fifteen-year-old students (see also Kombrink 1997). The researchers developed about 15 lessons devoted to descriptive statistics and exploratory data analysis. As part of the teaching experiment, students investigated data concerning leisure time and media, which had been collected from 120 fifteen-year-olds from their school.

Among the statistical notions and graphs taught were histograms, box plots, dot plots, mean, median, quartiles, range, and interquartile range. Students used the software WINSTAT. The experiment was informed by the exploratory study reported in Konold et al. (1997) and Biehler (1997). A major goal was for students to learn to interpret the box plot as an indicator of distribution shape and to correctly interpret smaller components of a box plot as indicating areas of higher density. As part of the instruction, students sketched box plots from histograms of the data, and likewise created histograms based on box plots. Students also learned to anticipate how skewed distributions would appear in histograms and box plots. With these kinds of exercises, they also learned to see the parts of the box plot as indicating different average densities.

What surprised the researchers was that students did not learn to interpret the parts of the box plot as measures of center or spread. They regarded the median primarily as a cut point and not as a way to summarize where the data were centered.[iii] Furthermore, the students interpreted the interquartile range as “the spread of the middle half of the data”, rather than as a measure of spread that is a property of the whole data set. For the difference of the median and the lower quartile, the students coined the phrase “spread of the second quarter”, and so on. Biehler (2001) called this interpretation of box plots the “shape summary interpretation”. This interpretation seems to be consistent with Tukey’s idea to introduce several “measures of spread” instead of summarizing spread by only one number such as the classic standard deviation. In retrospect, these results were generally consistent with the primary goal of the teaching experiment, which was to focus on the idea of distribution shape, leaving the idea of summarizing the distribution using various statistical measures as a next step.

The shape summary interpretation can be contrasted with a “center + spread interpretation”, in which the interquartile range is a measure of spread of the whole data set (Biehler, 1997). The interquartile range is introduced in many statistics textbooks as a measure of spread with two advantages over the standard deviation: it can be easily calculated and geometrically interpreted, and it is robust against outliers. From this perspective, box plots can be used for hierarchical group comparisons according to center, spread, shape, and outliers (Erickson & Nonsanchuk, 1977, emphasize this interpretation).

As stated earlier in this paper, box plots are especially useful for comparing two or more data sets. Using the center + spread perspective, we might compare two groups by looking first at the location of the centers of each, then at the spreads (IQRs), and finally at finer details of the two distributions. This strategy of box plot comparison concentrates on the location of the distribution and the “middle grey box” as measures of spread.

However, taking this perspective turned out to be more difficult for the students than the researchers had expected. In contrast, many students tended to compare all the five numbers—minimum, lower quartile, median, upper quartile, and maximum—regarding each with equal importance. When all the five numbers of the box plot of one group were higher than those of another group, the students would conclude that one group had “larger values” than the other (thus using a “shift model”). When these differences were not all in the same direction, they did not know what to conclude.

Based on these findings, Biehler (2001) made the following recommendations for future teaching experiments.

- We need to find contexts that more easily evoke a center + spread interpretation.
- Rather than describing the interquartile range only as the spread of the “middle half” of the data, we should introduce a second developmental step and encourage viewing the components as a measure of deviation. The *upper quartile - median* (Q_3-M) can be seen as an indicator of how far the higher values deviate from the median: it is mathematically equal to the median deviation of the higher values from the median. Similarly, the *median-lower quartile* (M-Q_1) is an indicator of how far the lower values deviate from the median. The interquartile range is then composed of these two parts and can be seen as a measure of spread of the whole data set. This sets the stage for later introducing measures of spread as mean or median deviations from a center (such as the variance and standard deviation).
- Box plot comparison should be guided by certain prototypical examples and concepts, such as the “shift model” and deviations from it.

Teaching experiments with older students (above 18) using this approach seemed to be effective. Whether it would work with younger students and in shorter interventions is unclear.

Conclusions

We have described in this paper various challenges involved in learning about box plots and have recommended that we not include the box plot in the statistics curricula until at least the high school level, unless we are willing to devote considerably more instructional time helping students learn to interpret median, quartiles, and box plots. In the four main arguments we highlighted the following issues.

A. Students are inclined to view data as individual cases whereas box plots only provide aggregate information. As preparation to learning box plots, we might productively use a combination of dot plots and quartiles (Figure 2).

B. Two of the difficulties of box plots are that they display densities rather than frequencies and that, somewhat counterintuitively, the relative density is negatively related to the size of a box-plot component.

C. Though the median is relatively easy to learn as a procedure for finding a middle value or as a cut point in a dot plot, it is more difficult to develop the view that it is a measure of center for the group.

D. Similarly, the IQR is hard to develop as a measure of spread. One of the difficulties with real data (especially if they are rounded to integers) is the occurrence of ties; another problem is the variety of definitions of quartiles. When using educational software, the teacher should certainly be aware of what definition the software uses.

Throughout the paper we have suggested several alternative ways of learning median, quartiles, and box plots. In our experience, these alternative approaches are effective, but they require more time than is typically available in middle school classrooms. It is largely for these reasons that we think it unwise to teach box plots to students younger than 14 or 15 years old, especially given the fact that many students of this age are still struggling with the meaning and usage of percentages.

The question of when to introduce box plots raises a more general question: What general principles do and should we use in determining which graphical representations we emphasize at various grades? We offer the following tentative list:

- how likely it is that students will encounter that representation in other disciplines or in the media,
- the difficulty they pose to students,
- how well they support reasoning about the deeper concepts targeted at that grade.

We hope that future research will provide information that will help educators decide the placement of various statistical representations and concepts in the various grades. In particular, we see the need for research that explores the kinds of tools and representations we can provide younger students that allow them to notice and quantify the amount of variability in a distribution of values and to come to see measures of variability and center as group characteristics.

References

Arbeitskreis Stochastik (2003). Empfehlungen zu Zielen und zur Gestaltung des Stochastikunderrichts. *Stochastik in der Schule*, 23(3), 21-26.

Bakker, A. (2004). *Design research in statistics education; On symbolizing and computer tools*. Utrecht, the Netherlands: CD Beta Press. Online: www.fi.uu.nl

Bakker, A., & Gravemeijer, K. P. E. (in press). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

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.

Biehler, R. (1982). *Explorative Datenanalyse - eine Untersuchung aus der Perspektive einer deskriptiv-empirischen Wissenschaftstheorie [Exploratory data analysis - an investigation from the perspective of a descriptive-empirical scientific theory]* (Vol. 24). Bielefeld: Universität Bielefeld.

Biehler, R. (1997). Students' difficulties in practising computer supported data analysis - Some hypothetical generalizations from results of two exploratory studies [http://www.dartmouth.edu/~chance/teaching_aids/IASE/14.Biehler.pdf]. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, the Netherlands: International Statistical Institute.

Biehler, R. (2001). Statistische Kompetenz von Schülerinnen und Schülern - Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen. In M. Borovcnik, J. Engel & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht* (pp. 97-114). Hildesheim, Germany: Franzbecker.

Biehler, R. (2003). *Interrelated learning and working environments for supporting the use of computer tools in introductory courses [http://www.ph-ludwigsburg.de/iase/proceedings/index.html#papers]*. Paper presented at the IASE Satellite conference on Teaching Statistics and the Internet, Berlin.

Biehler, R., & Kombrink, K. (1999). Der Datendetektiv auf Spurensuche. *Mathematik lehren, 97*, 23-46.

Biehler, R., & Steinbring, H. (1991). Entdeckende Statistik, Stengel-und-Blätter, Boxplots: Konzepte, Begründungen und Erfahrungen eines Unterrichtsversuches [Statistics by discovery, stem-and-leaf, boxplots: Basic conceptions, pedagogical rationale, and experiences from a teaching experiment]. *Der Mathematikunterricht, 37*(6), 5-32.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801-824.

Cobb, P., Gravemeijer, K. P. E., Bowers, J., & Doorman, M. (1997). *Statistical Minitools* [applets and applications]. Nashville & Utrecht: Vanderbilt University, TN & Freudenthal Institute, Utrecht University.

Cobb, P., McClain, K., & Gravemeijer, K. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction, 21,* 1-78.

David, H. A. (1998). Early sample measures of variability. *Statistical Science, 13*(4), 368-377.

Erickson, B. W., & Nonsanchuk, T. A. (1977). Understanding data. Toronto: McGraw-Hill.

Freund, J. E., & Perles, B. M. (1987). A new look at quartiles of ungrouped data. *The American Statistician, 41*(3), 200 - 203.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic enquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337-364.

Hoaglin, D. C. (1983). Letter values: A set of selected order statistics. In D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.), Understanding robust and exploratory data analysis (pp. 33-57). New York: Wiley.

Kombrink, K. (1997). *Kompetenzen von Schülerinnen und Schülern im Umgang mit Darstellungsmitteln und Begriffen der beschreibenden Statistik*. Bielefeld: Universität Bielefeld: Fakultät für Mathematik.

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., & Miller, C. (2004). Tinkerplots. Data analysis software for middle school curricula. San Francisco: Key Curriculum Press.

Konold, C., Pollatsek, A., & Well, A. D. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 151-167). Voorburg, the Netherlands: International Statistical Institute.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A. D., Wing, R., et al. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings* *of the International Conference on Teaching Statistics [CD-ROM], Cape Town*. Voorburg, the Netherlands: International Statistics Institute.

Landwehr, J. M., & Watkins, A. E. (1986). *Exploring data (and exploring data teacher's edition)*. Palo Alto, CA: Dale Seymour Publications.

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations in box plots. *The American Statistician, 32*(1), 12-16.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*, 20-39.

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

Scheaffer, R. L. (2000). Statistics for a new century. In M. J. Burke (Ed.), *Learning mathematics for a new century, 2000 Yearbook* (pp. 158-173). Reston, VA: NCTM.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

[i] In Figure 1, the fences are Q1 - 1.5*(Q3-Q1) and Q3 + 1.5*(Q3-Q1). Whiskers are drawn to the extreme values that are inside the fences. This adds a complexity, but in fact box plots without this feature are not so useful because they are not resistant against outliers.

[ii] For example, assume a data set has n values. Order the data according to values and calculate $n/4$. If this is a whole number, take the mean of the data at the positions $n/4$ and $n/4 + 1$. If it is a fraction, take the data value at the next whole number position. So if n = 23, n/4 = 5.75 we have to take the data value at position 6. If n is of the type $4k+1$ (e.g., 9, 13, 16, etc.) then this rule is equivalent to Tukey's rule; if $n = 4k +3$ this rule is equivalent to just omitting the median from both groups when calculating the quartile (as the teachers mentioned above did).

[iii] All authors (Bakker, 2004; Biehler, 1997; Konold et al., 1997) have observed that students prefer the mean over the median, because the mean uses all the data whereas the median does not change when some data are changed. Seventh graders in Bakker's study considered the mean "more precise".