ISSUE 14 AUGUST 2018

CAMBRIDGE Mathematics ESPRESSO

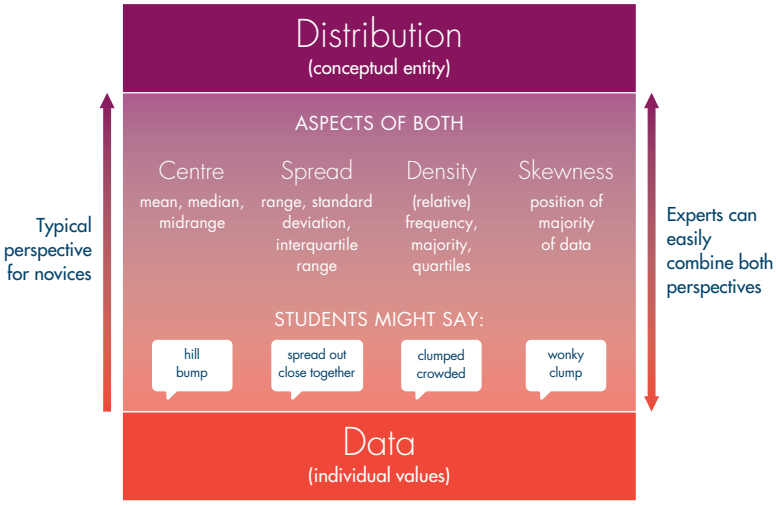
RESEARCH, FILTERED BY CAMBRIDGE MATHEMATICS

TALKING POINT:

WHAT DOES RESEARCH SUGGEST ABOUT EFFECTIVE WAYS TO INTRODUCE COMPARISON BETWEEN DATA SETS?

IN SUMMARY

- Comparison between data sets is a crucial composite skill and more research/teacher support is required in this area
- Students should be encouraged to explore and inquire about data and to create and use multiple representations
- Students should be encouraged to start with informal ideas about shape before moving on to more formal conceptions (and language) such as mean, median and, later, quartiles
- Teachers should help students to understand data flexibly: as an aggregate as well as a set of individual cases (see infographic)
- From early stages teachers should be clear with students about the difference between sample and population, and explore the effect of sampling in more detail later on
- Students should start with identifying and then interpreting differences between data sets, considering ideas of *measurable conjectures, variability, context* and *inference*
- Specifically designed software such as TinkerPlots and Fathom can support learning



A structure to show the relation between data and distribution (adapted from Bakker & Gravemeijer, 2004)

1

Comparing groups of data (and asking whether the differences are meaningful[1]) is a critical skill[2]: it is at the 'heart of statistics' because it encompasses and uses ideas such as data, aggregation, variation and distribution[3] and has been described as a crucial 'organising conceptual structure'[4]. There is some consensus that students should start by developing their intuitive understanding, with a focus on informal characteristics and language (i.e. the shapes made by the data) before moving to more formal ideas of spread and centre (to include means, medians and quartiles)[11]. However, this is a relatively recent[5] and sparse[2] research area and as such there is not yet an adequate or standardized set of language to help describe or teach it[2,6] and teachers need better resources and support for conceptual understanding[6].

IMPLICATIONS: Comparing two data sets is a crucial skill in statistics learning that draws together many other important ideas

Students should start by developing intuitive and informal ideas about shapes made by the data before moving onto formal measures of spread and centre

There is a need for more research, resources and support for teachers in this area

'It is important that in lessons and activities designed to develop inferential reasoning, students remain focused on the big picture, rather than getting distracted by activities like constructing graphs by hand or calculating statistics'
Hornby & Macey, 2018

'The notion that one can use available data to make claims about the unknown is the most powerful tool in statistics; one which we believe should be accessible earlier'
Makar et al, 2011

CAMBRIDGE UNIVERSITY PRESS
UNIVERSITY OF CAMBRIDGE Faculty of Mathematics
UNIVERSITY OF CAMBRIDGE Faculty of Education
Cambridge Assessment

2

Comparing data sets can be done at different ages[2] and should start informally by focusing on ideas of shape[3], allowing students to explore visual ideas such as gaps, clusters, hills and bumps[1,3]. It has been argued that comparing data sets is the first step towards developing students' beginning notions of inference[7]. It is better for younger students to compare two groups with one another than one group against a hypothetical model[4]. Some research findings suggest students should consider sampling quite early on[3]; others advise leaving more complex ideas about sample size until later[6]. Early clarity is suggested about the distinction between *description* when comparing populations and *inference* when working with samples[8].

IMPLICATIONS: Even young students (primary age) can start to develop informal ideas of data set comparison by exploring visual ideas of shape

Detailed explanation of sample size issues may be best left until later but early introduction of populations and samples is suggested

3

One of the biggest issues surrounding data comparison is the difficulty students have in conceiving of data sets as an aggregate, not just individual values[1,4,11,] which is also exemplified in the infographic as moving fluidly upwards and downwards between ideas and aspects of *distribution* and *data*. One way to support students to do this is to adopt a 'data detective' approach (considering not just 'how but also..why we collect and explore data'[9]). Even students who are familiar with formal measures such as mean and median often don't think to use them for comparing[1].

IMPLICATIONS: Teachers should pay careful attention to moving students towards a flexible understanding of data sets as an aggregate as well as individual values

After learning how to calculate the mean and median, students should be encouraged to use these as part of comparing data sets

An early focus on deep inquiry is suggested as effective

4

Specially designed software such as Fathom or TinkerPlots can support development in this area; some recent research has focused on ways to use these effectively[4]. This can allow learners to easily draw and investigate multiple – possibly simultaneous – representations such as box plots and hat plots (described as 'proto-box plots') which may show data in a way that is more intuitive for students[10] and help them to move towards an aggregate view of the data.

IMPLICATIONS: TinkerPlots and Fathom are well-researched software programs to support data set comparison

In particular, the use of statistical software may help students in drawing and interpreting box plots and hat plots, which may help learners move towards an aggregate view of data

Lucy Rycroft-Smith & Darren Macey, 2018

REFERENCES

1 Makar, K., & Confrey, J. (2002), Comparing two distributions: investigating secondary teachers' statistical thinking, *ICOTS-6 Papers for School Teachers,* Cape Town: International Association for Statistics Education

2 Schnell, S., & Büscher, C. (2015), Individual concepts of students comparing distributions, in K. Krainer & N. Vondrová (Eds.), *CERME 9 – Ninth Congress of the European Society for Research in Mathematics Education* (pp. 754–760), Prague, Czech Republic: Charles University in Prague, Faculty of Education and ERME.

3 Ben-Zvi, D., Makar, K. and Garfield, J. (eds) (2018) International Handbook of Research in Statistics Education, Chapter 5.5, pp163–169, Springer International Publishing

4 Bakker, A. (2004), Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64–83.

5 Pfannkuch, M., Budgett, S., Parsonage, R., & Horring, J. (2004) Comparison of data plots: building a pedagogical framework, paper presented at the International Association for Statistics Education 2004, Lind, Sweden: 2004

6 Leavy, A. (2006) Using data comparison to support a focus on distribution: examining preservice teachers' understandings of distributions when engaged in statistical inquiry, *Statistics Education Research Journal*, v5 n2 p89–114

7 Watson, J. M., & Moritz, J. B. (1998), The Beginning of Statistical Inference: Comparing Two Data Sets. *Educational Studies in Mathematics*, *37*(2), 145–168.

8 Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011) Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(2), 247–295.

9 Konold, C., & Pollatsek, A. (2002), Data analysis as the search for signals in noisy processes. Journal for Research in Mathematics Education, 33(4), 259-289.

10 Watson, J. M., Fitzallen, N., Wilson, K. G., & Creed, J. F. (2008),The Representational Value of Hats, *Mathematics Teaching in the Middle School*, *14*(1).

CAMBRIDGE UNIVERSITY PRESS

UNIVERSITY OF CAMBRIDGE Faculty of Mathematics

UNIVERSITY OF CAMBRIDGE Faculty of Education

Cambridge Assessment