Susan Jo Russell and Jan Mokros

What Do Children Understand about Average?

The statistical idea we come across most frequently is the idea of average. Children in fourth grade and beyond fairly easily learn to apply the algorithm for finding the mean, but what do they understand about the mean as a statistical idea?

Many students do not have opportunities to learn about various kinds of averages as statistical concepts. They view an average as a number found by a particular procedure rather than as a number that represents and summarizes a set of data. Students may learn to find a mode, median, or mean—which technically are all averages even though average often refers to the mean—but they do not necessarily know how these statistics relate to the data being represented.

Average as a Statistical Idea

To investigate students' understanding of the idea of average, we use what we call "construction" problems. Instead of asking students to find the average for a given set of numbers, we give students an average and ask them what could be the data set it represents. This kind of problem is similar to situations we often encounter in life; we read about a median or mean in the newspaper or come across it in our work and need to interpret what might be represented by that number. Statistically literate readers can think about such a statement as "the median price of a house is $150 000" or "the average family size is 3.2" in terms of what it tells or does not tell them about the distribution of the data.

Most students are unable to imagine what kind of data an average might represent. One student, who had had plenty of experience calculating the mean, said, "I know how to get an average, but I don't know how to get the numbers to go into an average, from an average."

Asking students to imagine what the data could be for a given average yields interesting insights into students' thinking about the relationship between data and the average of the data. They do not find it easy to use the memorized algorithm to construct data. They have to think about how the average represents the data. You might want to try the following problem, one we gave to students, before reading about how students solve it.

> We took a survey of the prices of nine different brands of potato chips. For the same-sized bag, the typical or usual or average price for all brands was $1.38. What could the prices of the nine different brands be?

We use the language "typical or usual or average" to keep the conversation open to any ways that students have to think about an average. When they show us one way, we ask them for other ways so that we get a view of the range of their thinking. Since these problems are administered in an interview, we are able to interact, ask questions, and probe students' thinking.

Average as mode

In interviews with fourth graders, many students consistently associated the "typical or usual or average" value with the mode. In construction problems, they produced the data set by making all or most of the values the same as the average value. They might have made a few adjustments to the data when pushed, but despite probing for other approaches, these students stuck to a view of the average as the most frequent piece of data. As one fourth grader explained, "Okay, first, not all chips are the same, as you told me, but the lowest chips I ever saw was $1.30 myself, so since the typical price is $1.38, I just put most of them at $1.38, just to make it typical, and highered the prices on a couple of them, just to make it realistic."

Average as median

Another group of students relied more on reasonableness in constructing the data from the average. They drew on what was realistic in their own lives

Susan Jo Russell directs a K–5 curriculum project called Investigations in Number, Data, and Space. Her work focuses on how practicing teachers can learn more about mathematics and children's mathematical thinking. Jan Mokros codirects TERC's Math Center and recently wrote a book for parents called Beyond Facts and Flashcards: Exploring Math with Your Kids *(Portsmouth, N.H.: Heinemann Publishers).*

Edited by **Donald L. Chambers,** *Wisconsin Center for Education Research, University of Wisconsin—Madison, Madison, WI 53706*

The action-research ideas in this article were prepared by Donald Chambers.

but were also concerned about mathematical reasonableness. These students usually thought about an average as being a middle value and constructed their data sets so that high values were countered by low values. They did not necessarily find a precise middle but rather centered their average value roughly in the middle of their data.

Some students used an even stronger understanding of "middle." These students often developed perfectly symmetrical distributions around their average value. For example, in the potato-chip problem they would put one price at $1.38, then one at $1.37 and one at $1.39, then one at $1.36 and one at $1.40, then perhaps one at $1.30 and one at $1.46, and so forth. If we moved one of the data points on their distribution to a new value, they could easily adjust their data set to keep the same average. However, if we introduced a constraint that did not allow them to make a perfectly symmetrical distribution, they were often stumped. When we asked students to make prices for the potato-chip problem without using the value $1.38, most said that the task could not be done. Others made minor adjustments to their symmetrical distribution, such as changing the $1.38 to $1.37 and arguing that this modification was the best they could do.

To help students learn more about averages, it makes sense to build on their developing understandings. Students often invent the idea of looking at the middle of the data as a way to compare data sets. The median—the value of the middle piece of data—connects clearly to these informal understandings and has increasingly become that statistic of choice in many real data-analysis contexts because unusually high or low values in the data do not greatly affect it. Although developing an understanding of how the median represents the data is not without complexity (see Russell and Corwin [1989, 54]), it is a good place for upper elementary students to begin learning about averages.

When finding a median, we can actually point to a value in the data set—the middle value or the midpoint between the two middle values. Finding the middle value is easy if all the data are arranged in order. For example, to find the median of the heights of the students in a classroom, the students could line up in order of height. The height of the middle person, or the midpoint between the heights of the two middle people, is the median value. Half of the data are below the median, and half of the data are above it.

With experience, students can begin envisioning the variety of data sets that might be represented by a median value of, for example, 48 inches. **Figure 1** pictures three such data sets: (a) is fairly symmetrical around 48 but has a small range; (b) is also fairly symmetrical but has a big range; and (c) is rather bimodal, with few data near the median itself.

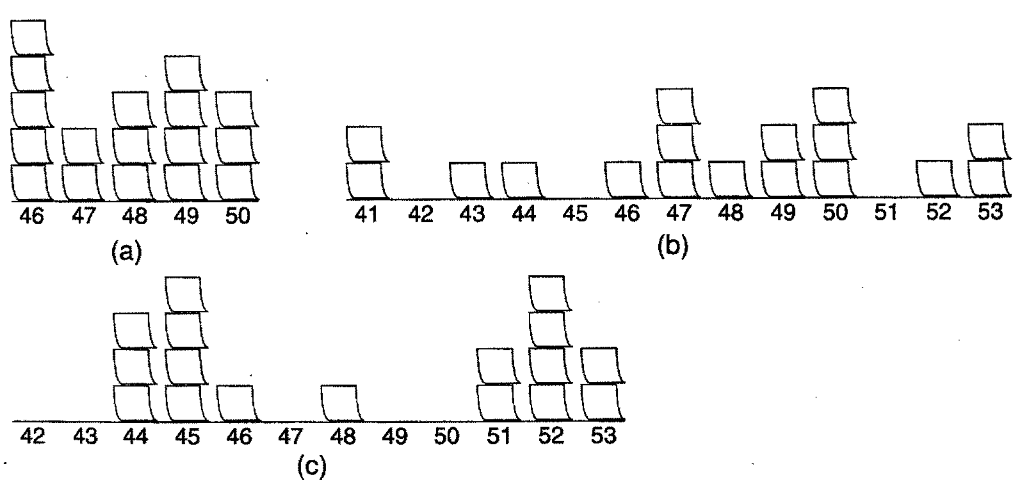
Average as a procedure

Fourth graders who tried to use the algorithm for the mean to solve the potato-chip problem generally got stuck and frustrated. One student divided $1.38 by 9, resulting in a price close to $0.15. We asked her if pricing the bags at $0.15 would result in a typical price of $1.38, and she responded, "Yeah, that's close enough." Another student chose for her prices pairs of numbers that totaled $2.38, such as $1.08 and $1.30. She thought that this method resulted in an average of $1.38. Other students did other meaningless calculations, such as choosing one price then getting the next price by subtracting $0.38.

Through many interviews, we have found that most students' knowledge of the procedure for finding a mean is not at all connected to understanding what the mean represents. However, we have also found that fourth through sixth graders are

Students cannot imagine the data an average represents

FIGURE 1 Three data sets, each with a median of 48



Using one set of poorly understood ideas may not foster understanding of related ideas

developing important understandings about the nature of an average, including these:

- A value in the middle of the data can represent a set of data.
- The average is situated in the data in such a way that values higher than the average are countered by values lower than the average.
- An average for a given set of data furnishes a reasonable sense of the values of the data.
- The shape of the data represented by an average should reasonably reflect what we know about the context from which these values are derived, for example, heights or family size.

Learning about the Mean

Students need many experiences with data sets and the median before they can understand how the mean represents the data. Unlike the median, the mean is a mathematical abstraction. We can "see" where the median falls among the data. The mean has no such clear identity within the data themselves; its value may not appear in the data set at all. The average family size may be a number such as 3.6, a number that is not the size of any family. The mean is a mathematical construction that represents certain relationships in the data—a kind of abstraction that may be new to many upper elementary school students.

Over the years mathematics educators have tried to develop models that support students as they build an understanding of the mean. Early in our work with sixth graders, we assumed that students would most effectively learn this relationship by having an opportunity to transform a set of data into "fair shares." If each person has a certain number of pets, the way to find the average is to pool the pets then distribute them equally among the children who pooled their pets. This model reflects exactly what happens when the averaging algorithm is used: the data are pooled then divided evenly, as if each piece of data had the same value. That value is the mean.

We discovered that although this model is effective in teaching children about division, it simply does not help them think about the statistical relationship between the data and the mean. The problem with this model is that in the act of redistributing quantities so that each data point has the same value, the relationship between the actual data and the mean is completely obscured.

For example, consider a problem in which eight children checked the following numbers of books out of the library: 5, 4, 3, 2, 1, 4, 5, and 2. What is the mean number of books checked out by each child? In the fair-shares model all the books are put together then the twenty-four books are redistributed into eight groups (or divide by 8) so that three books are in each group. We then have an image of eight children with three books each. The mean is 3. But what is the relationship of this "3" to the original set of data? In the course of redistributing the data, we have lost the original set of data. It is no wonder, then, that children cannot see the relationship between data and mean using this model. They do not have the opportunity to see them together! Once the total is calculated, the individual data are lost. This scenario is not what happens in real life, where averaging rarely involves making fair shares. In life, and in real statistical situations, the data still exist in their original form. We found that learning the fair-shares model to find the mean did not result in better understanding of the relationship between mean and data. Students were still unable to solve construction problems, such as the potato-chip problem described in the foregoing, in which their job was to create a data distribution with a particular mean value.

Another model emphasizes the mean as the value about which all the data "balance." Rather than have the same number of data points on either side, as with the median, we want the sums of the differences from the mean on either side to be equal. **Table 1** shows two distributions for which the mean value is \$1.38. In both distributions, the sum of the differences from the mean is 0. In our research, very few students—and none younger than sixth grade—were developing notions about the mean as a point of balance.

TABLE 1

Two distributions for which the mean value is \$1.38

Case 1		Case 2	
Price	Price−\$1.38	Price	Price−\$1.38
\$1.40	+0.02	\$1.50	+0.12
\$1.39	+0.01	\$1.49	+0.11
\$1.38	0.00	\$1.39	+0.01
\$1.37	−0.01	\$1.34	−0.04
\$1.36	−0.02	\$1.18	−0.20

Balance models are not new (Pollatsek, Lima, and Well 1981) but can be problematic because they depend on an understanding of the relationship between weight and distance on a balance beam. Using one poorly understood set of ideas—the physical relationship of weight and distance—may not help students understand another set of difficult ideas—the numerical relationship between the mean and the data.

The unpacking model

In our work with sixth graders, we use a new model