Data Analysis and Probability

STANDARD for Grades 3–5

Instructional programs from prekindergarten through grade 12 should enable all students to—

Expectations

In grades 3–5 all students should–

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them	• design investigations to address a question and consider how data-collection methods affect the nature of the data set; • collect data using observations, surveys, and experiments; • represent data using tables and graphs such as line plots, bar graphs, and line graphs; • recognize the differences in representing categorical and numerical data.
Select and use appropriate statistical methods to analyze data	• describe the shape and important features of a set of data and compare related data sets, with an emphasis on how the data are distributed; • use measures of center, focusing on the median, and understand what each does and does not indicate about the data set; • compare different representations of the same data and evaluate how well each representation shows important aspects of the data.
Develop and evaluate inferences and predictions that are based on data	• propose and justify conclusions and predictions that are based on data and design studies to further investigate the conclusions or predictions.
Understand and apply basic concepts of probability	• describe events as likely or unlikely and discuss the degree of likelihood using such words as *certain*, *equally likely*, and *impossible*; • predict the probability of outcomes of simple experiments and test the predictions; • understand that the measure of the likelihood of an event can be represented by a number from 0 to 1.

Data Analysis and Probability

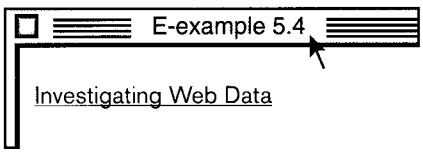
In prekindergarten through grade 2, students will have learned that data can give them information about aspects of their world. They should know how to organize and represent data sets and be able to notice individual aspects of the data—where their own data are on the graph, for instance, or what value occurs most frequently in the data set. In grades 3–5, students should move toward seeing a set of data as a whole, describing its shape, and using statistical characteristics of the data such as range and measures of center to compare data sets. Much of this work emphasizes the comparison of related data sets. As students learn to describe the similarities and differences between data sets, they will have an opportunity to develop clear descriptions of the data and to formulate conclusions and arguments based on the data. They should consider how the data sets they collect are samples from larger populations and should learn how to use language and symbols to describe simple situations involving probability.

Investigations involving data should happen frequently during grades 3–5. These can range from quick class surveys to projects that take several days. Frequent work with brief surveys (How many brothers and sisters do people in our class have? What's the farthest you have ever been from home?) can acquaint students with particular aspects of collecting, representing, summarizing, comparing, and interpreting data. More extended projects can engage students in a cycle of data analysis—formulating questions, collecting and representing the data, and considering whether their data are giving them the information they need to answer their question. Students in these grades are also becoming more aware of the world beyond themselves and are ready to address some questions that have the potential to influence decisions. For example, one class that studied playground injuries at their school gathered evidence that led to the conclusion that the bars on one piece of playground equipment were too large for the hands of most students below third grade. This finding resulted in a new policy for playground safety.

Investigations involving data should happen frequently during grades 3–5.

Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

At these grade levels, students should pose questions about themselves and their environment, issues in their school or community, and content they are studying in different subject areas: How do fourth graders spend their time after school? Do automobiles stop at the stop signs in our neighborhood? How can the amount of water used for common daily activities be decreased? Once a question is posed, students can develop a plan to collect information to address the question. They may collect their own data, use data already collected by their school or town, or use other existing data sets such as the census or weather data accessible on the Internet to examine particular questions. If students collect their own data, they need to decide whether it is appropriate to conduct a survey or to use observations or measurements. As part of their plan, they often need to refine their question and to consider aspects of data collection such as how to word questions, whom to ask, what and when



to observe, what and how to measure, and how to record their data. When they use existing data, they still need to consider and evaluate the ways in which the data were collected.

Students should become familiar with a variety of representations such as tables, line plots, bar graphs, and line graphs by creating them, watching their teacher create them, and observing those representations found in their environment (e.g., in newspapers, on cereal boxes, etc.). In order to select and interpret appropriate representations, students in grades 3–5 need to understand the nature of different kinds of data: categorical data (data that can be categorized, such as types of lunch foods) and numerical data (data that can be ordered numerically, such as heights of students in a class). Students should examine classifications of categorical data that produce different views. For example, in a study of which cafeteria foods are eaten and which are thrown out, different classifications of the types of foods may highlight different aspects of the data.

As students construct graphs of ordered numerical data, teachers need to help them understand what the values along the horizontal and vertical axes represent. Using experience with a variety of graphs, teachers should make sure that students encounter and discuss issues such as why the scale on the horizontal axis needs to include values that are not in the data set and how to represent zero on a graph. Students should also use computer software that helps them organize and represent their data, including graphing software and spreadsheets. Spreadsheets allow students to organize and order a large set of data and create a variety of graphs (see fig. 5.20).

Fig. **5.20.**
Spreadsheet with weather data

	A	B	C	D
1	Daily Precipitation and Temperatures for San Francisco, California			
2				
3		Precipitation (inches)	Temperature (°F)	
4	Date		Hi	Low
5	1/1	0.01	58	48
6	1/2	0.88	60	51
7	1/3	0.43	58	50
8	1/4	0.25	56	44
9	1/5	0	51	40
10	1/6	0.25	54	40
11	1/7	0.09	50	47
12	1/8	0	51	47

When students are ready to present their data to an audience, they need to consider aspects of their representations that will help people understand them: the type of representation they choose, the scales used in a graph, and headings and titles. Comparing different representations helps students learn to evaluate how well important aspects of the data are shown.

Select and use appropriate statistical methods to analyze data

In prekindergarten through grade 2, students are often most interested in individual pieces of data, especially their own, or which value is "the most" on a graph. A reasonable objective for upper elementary and middle-grades students is that they begin to regard a set of data as a whole that can be described as a set and compared to other data sets (Konold forthcoming). As students examine a set of ordered numerical data, teachers should help them learn to pay attention to important characteristics of the data set: where data are concentrated or clumped, values for which there are no data, or data points that appear to have unusual values. For example, in figure 5.21 consider the line plot of the heights of fast-growing plants grown in a fourth-grade classroom (adapted from Clement et al. [1997, p. 10]). Students describing these data might mention that the shortest plant measures about 14 centimeters and the tallest plant about 41 centimeters; most of the data are concentrated from 20 to 23 centimeters; and the plant that grew to a height of 41 centimeters is very unusual (an outlier), far removed from the rest of the data. As teachers guide students to focus on the shape of the data and how the data are spread across the range of values, the students should learn statistical terms such as *range* and *outlier* that help them describe the set of data.

Plant Height Data	
Height (in cm)	Number of Plants
15	|
20	|||\
22	||
23	𝍸 ||
40	|

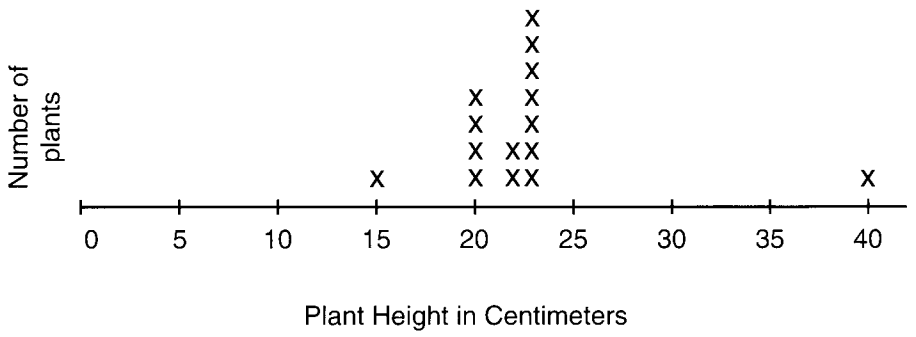


Fig. **5.21.**
Plant height data from a fourth-grade class

Much of students' work with data in grades 3–5 should involve comparing related data sets. Noting the similarities and differences between two data sets requires students to become more precise in their descriptions of the data. In this context, students gradually develop the idea of a "typical," or average, value. Building on their informal understanding of "the most" and "the middle," students can learn about three measures of center—mode, median, and, informally, the mean. Students need to learn more than simply how to identify the mode or median in a data set. They need to build an understanding of what, for example, the median tells them about the data, and they need to see this value in the context of other characteristics of the data. Figure 5.22 shows the results of plant growth in a third-grade classroom (adapted from Clement et al. [1997, p. 10]). Students should compare the two sets of data from the fourth- and third-grade classrooms. They may note that the median of the fourth-grade data is 23 centimeters and the median of the third-grade data is 28 centimeters. This comparison provides information