

## EXTERNAL REPRESENTATIONS FOR DATA DISTRIBUTIONS: IN SEARCH OF COGNITIVE FIT

STEPHANIE LEM

*KU Leuven*

*stephanie.lem@ppw.kuleuven.be*

PATRICK ONGHENA

*KU Leuven*

*patrick.onghena@ppw.kuleuven.be*

LIEVEN VERSCHAFFEL

*KU Leuven*

*lieven.verschaffel@ppw.kuleuven.be*

WIM VAN DOOREN

*KU Leuven*

*wim.vandooren@ppw.kuleuven.be*

### ABSTRACT

*Data distributions can be represented using different external representations, such as histograms and boxplots. Although the role of external representations has been extensively studied in mathematics, this is less the case in statistics. This study helps to fill this gap by systematically varying the representation that accompanies a task between participants, and assessing how university students use such representations in comparing aspects of data distributions. Following a cognitive fit approach, we searched for matches between items and representations. Depending on the item, some representations led to better achievement than other representations. However, due to the low overall accuracy rates and various difficulties that students displayed in interpreting these representations, we cannot make strong claims regarding matches between items and representations.*

**Keywords:** *Statistics education research; Cognitive fit theory; Boxplots; Histograms; Descriptive statistics*

Distribution is one of the key concepts of statistics. Distributions, which can be empirical or theoretical, form the basis for statistical decision making. In this article we focus on empirical (or data) distributions. Understanding such distributions is considered a prerequisite for understanding more complex distributions, such as sampling distributions (e.g., Meletiou & Lee, 2002). Although data distributions are accessible only through external representations, the research literature focusing on reasoning about distribution so far has not paid much explicit attention to the role of the actual representations that are used to present a data set to participants (e.g., Brase, 2009). In other areas of the mathematics and science education literature, however, it has been stressed that external representations have an important impact on students' learning and problem solving (e.g., Ainsworth, Bibby, & Wood, 1998; Goldin, 2002; Kaput, 1992). In this article we report a study that focused on students' use of four widely used external representations to compare two distributions with regard to mean, median, variation, and skewness: the dotplot, the (grouped) histogram, the boxplot, and descriptive statistics. In previous studies regarding distributional reasoning typically only one specific representation was used for each item (e.g., Meletiou & Lee; Pfannkuch, 2006; Prodromou & Pratt, 2006; Watson & Moritz, 1999), whereas in the current study we systematically varied the representation that accompanied an item between participants. This enabled us to study the actual

influence of external representations on distributional problem solving. Taking a cognitive fit perspective, we tried to find matches between the four kinds of external representations and the four types of items (i.e., comparing two distributions with respect to their mean, median, variation, or skewness).

## 1. THEORETICAL AND EMPIRICAL BACKGROUND

The theoretical and empirical background is divided into three sections. First, we define distributional reasoning, discuss why it is an important competence for students, and review the available empirical research. The second section discusses the scarce empirical evidence related to students' use of external representations in distributional reasoning, and relates this evidence to theoretical principles on the design of graphs. In the third section we take a cognitive fit approach to problem solving with external representations, and attempt to find matches between items and external representations.

### 1.1. DISTRIBUTIONS

Leavy (2006) defined the concept of distribution as “the arrangement of values of a variable along a scale of measurement resulting in a representation of the observed or theoretical frequency of an event” (p. 90). Three different types of distributions can be distinguished: empirical or data distributions, sampling distributions, and theoretical distributions. Statistical tests are based on the last two types of distributions, while empirical distributions have an important role in descriptive statistics (Moore, 1990). Wild (2006) emphasized the importance of being able to see patterns in data distributions. This means that, in order to look at data as a distribution, one also needs to look at features of the whole dataset (i.e., take a “global” view on data) instead of merely looking at features of individual values (i.e., take a “local” view of data). Konold, Higgins, Russell, and Khalil (2004) proposed six different lenses through which one can view data. These lenses range from looking at individual values of observations to looking at data as an aggregate. They stated, however, that we should not see these different lenses as separate levels in a hierarchy of reasoning, with a higher level being superior to a lower, as each lens can be useful depending on the specific item, data, and context.

Being able to reason about data distributions is considered to be an important outcome of statistics education (e.g., Bakker & Gravemeijer, 2004; Garfield & Ben-Zvi, 2004). It is an important part of data *description* and thus of data *analysis* (Moore, 1990). Furthermore, understanding data distributions is believed to help the individual in grasping more complex concepts of theoretical and sampling distributions (Meletiou & Lee, 2002). However, interpreting data distributions is for many students already a difficult skill to master (Garfield & Ben-Zvi, 2008).

A fair number of studies into distributional reasoning have already been conducted. The majority of these studies are qualitative, providing rich and detailed descriptions of the reasoning of a limited number of participants (e.g., Ben-Zvi, 2004; Konold, Pollatsek, Well, & Gagnon, 1997; Pfannkuch, 2006). The most frequently reported difficulty with reasoning about distributions is that many students fail to employ a global view on data (e.g., Bakker & Gravemeijer, 2004; Ben-Zvi & Arcavi, 2001; Garfield & Ben-Zvi, 2008). Instead, they focus only on (groups of) individual data points, an approach which is often referred to as a “local” view on data (e.g., Bakker & Gravemeijer; Konold et al., 2004). One consequence of this local reasoning is that students tend to focus on absolute frequencies rather than on proportions or chances (Konold et al., 1997). Other consequences include giving too much weight to outlying observations (Konold & Pollatsek, 2002), seeing measures of centre as properties of a particular value or observation (e.g., “That boy has an average height”) instead of viewing them as characteristics of the group of data (e.g., “The average height of pupils in this class is 1.60m”; Bakker, Biehler, & Konold, 2005), and relying completely on one single measure of centre or variability to come to a conclusion (e.g., only taking into account the mean of a skewed distribution with a large range; Kelly, Sloane, & Whittaker, 1997; Leavy, 2004).

## 1.2. EXTERNAL REPRESENTATIONS FOR DATA DISTRIBUTIONS

External representations are all inscriptions outside the mind that represent or stand for an object, state of affairs, or phenomenon (De Vries, Demetriadis, & Ainsworth, 2009). According to Schnotz (2005), two types of external representations can be distinguished: descriptive and depictive representations. Where descriptive or textual representations consist of symbols that have no visuospatial similarity with the represented object or event, depictive or graphical representations consist of icons that somehow resemble the represented object or event and use spatial dimensions. As both types of representations are very different, the way they are processed is also different, according to the integrative model of text and picture comprehension of Schnotz and Bannert (2003). Data distributions are, as stated above, only accessible by means of external representations. Therefore, being able to use these representations is very important. However, these representations can take a large number of forms, and can be both depictive and descriptive. Depictive representations of distributions that are widely used are dotplots, histograms, and boxplots. There are also descriptive ways of representing a distribution. One of them is to provide the numerical values for one or several descriptive statistics that summarise one or more characteristics of the distribution, thereby representing the distribution of a series of observed measurements. But even the listing of raw data by enumerating the observed values can be considered to be an external representation, as the raw data represent a series of measurements of an observed phenomenon.

As stated above, empirical evidence on the role of external representations in distributional problem solving is scarce. A systematic search of the literature yielded only a few studies that focused on the difficulties students have when interpreting external representations of distributions, and some other studies that provided mainly anecdotal evidence. However, none of them has ever explicitly compared these distributions. Hereafter we first discuss students' misinterpretation of graphical representations and attempt to interpret these misinterpretations in terms of the graph design principles developed by Tufte (1983) and Tversky (1997). Afterwards, we focus on the misinterpretation of non-graphical representations of data distributions.

Concerning histograms, the literature reports five interpretation problems. First, students mistakenly use the height differences between the bars of histograms as an indicator of variation, instead of using the spread and overall shape of the figure (Baker, Corbett, & Koedinger, 2002; Cooper & Shore, 2008; Meletiou & Lee, 2002). Second, students tend to confuse histograms with bar graphs, thinking that the height of a bar in a histogram represents the value of that bar, instead of the frequency or proportion (Baker et al., 2002; Bakker et al., 2005; Cohen, Smith, Chechile, Burns, & Tsai, 1996; Cooper & Shore, 2008; delMas, Garfield, & Ooms, 2005). Third, students sometimes interpret the horizontal axis as a time scale (delMas et al.). Fourth, they also misinterpret the classes of histograms, creating difficulties with the reading of frequencies of specific groups or values (delMas et al.). Finally, Watson and Moritz (1998) reported how grade 3–9 students thought that when one of two ungrouped histograms has a mode with a higher frequency, this implies that it has a higher mean than the other.

Concerning boxplots, Bakker et al. (2005) found three ways students misinterpret boxplots. First, students tend to misinterpret the area of the box as representing frequency or proportion. Second, students are confused by the various ways to deal with even numbers of data points when dividing data into four equal-sized groups. Finally, students find it difficult to see the median as descriptor of the distribution; they more readily perceive measures of centre as a characteristic of a specific value or observation. However, the third problem also occurs with other representations, suggesting it is not a difficulty specific to boxplots, but rather a consequence of the local view of data many students take.

With respect to dotplots our literature review did not yield any specific misinterpretations.

Possible explanations for the many difficulties students have interpreting graphical representations for data distributions can be found in design principles for graphs that have been proposed in the literature. Tufte (1983) proposed two measures to assess the quality of graphics: a data-ink ratio and a data-density index. Tufte's data-ink ratio is "the proportion of the graphic's ink devoted to non-redundant display of data-information" (p. 93). The higher this value, the better, as this way the eye is not distracted by ink that does not convey necessary information. The data-density index is "the size of the graphic in relation to the amount of data displayed" (p. 162). Again, the

higher this value the more information the graphic conveys, and the more effective and comparative our eye can be. We can use Tufte's data-ink ratio and data-density index to predict that graphical representations such as histograms or dotplots will be easier to interpret than boxplots: Boxplots are two-dimensional representations that require relatively much space to represent only five summary statistics and the distance among them, which could have been represented on a single dimension. Histograms and dotplots, on the other hand, present many more or even all available data points using about the same amount of space as a boxplot. This would make boxplots less effective according to Tufte's principles.

Tversky (1997) posed two important design principles, based on the way people interact with their environment: spatial metaphors and directionality. First, she stated that space in graphs should be used in a natural way. This means that elements in a graph that touch one another should also in reality be more related than elements that are placed further apart, and that area in the graph should correspond to the frequency or proportion of particular observations. Applying this idea to dotplots, we see that they consist of various separated dots. This can lead to local reasoning, because the space between the dots may lead one to believe that the dots are not related, whereas they are all part of the same distribution. On the other hand, the area use (stacks of dots) in a dotplot does correspond to the frequency of observations. This is in contrast with boxplots; whereas the area of a boxplot does not represent proportion or frequency, it is natural for people to interpret areas of different parts of a boxplot as representing proportions or frequencies of observations.

Second, Tversky stated that people tend to focus more on the vertical than on the horizontal dimension of a graph, which has also been shown empirically (e.g., Franklin & Tversky, 1990). This focus on verticality is also natural; people walk upright, grow in the vertical direction, and in language "higher" is associated with "better" or "more." This way, we can predict that people think of the vertical dimension of graphs as being very important, while giving less attention to the horizontal dimension. As a result, problem solvers may, for instance, correctly focus on the height of bars in histograms and stacks of dots in dotplots, but also incorrectly focus on the height differences in the bars of histograms when trying to assess variation. Both of Tversky's design principles can account for students' difficulties in interpreting graphical representations of data distributions which we discussed earlier, for instance in the finding of Watson and Moritz (1999) that students tend to focus on the frequency of the mode in two histograms when they wish to compare the means, believing the histogram with the higher modal bar will have the highest mean.

Although the design principles proposed by Tufte and Tversky are widely known and recommended (e.g., Kosslyn, 2006; Robbins, 2005), we are not aware of any empirical research that investigated the effect of the actual design of external representations on students' reasoning in regard to data distributions.

From the studies cited above, it becomes clear that several graphical aspects of representations can lead to misinterpretations. But also when students are confronted with a non-graphical representation of a distribution, such as the numerical value of one or more descriptive statistics, they can be misled. With respect to these numerical representations, students' main problem is that often they do not take into account more than one statistic to describe or compare distributions (Kelly et al., 1997; Leavy, 2004). This can be seen as a consequence of a local view on distributions.

### **1.3. COGNITIVE FIT THEORY**

In the previous subsection, we saw that graphical features of representations can be misleading. Non-graphical representations, such as descriptive statistics, do not have these misleading features. Still, graphical representations are often used as they allow us to see a distribution in one glance. Graphical representations can contribute to our understanding of a given distribution, conditional on the fact that they are designed well and fit with the features of the problem we are trying to solve. The design principles of Tufte (1983) and Tversky (1997) described in the previous subsection only prescribe in general terms how graphical representations should be designed. They do not take into account characteristics of the task that has to be solved by means of this representation, and hence do not help us in deciding which representations are helpful for solving which tasks. Cognitive fit theory (introduced by Vessey, 1991) is a framework that does take into account both characteristics of the representation and characteristics of the task, and, in doing so, may help to explain why reasoning

about a certain aspect of a distribution can be easier or more difficult, depending on the representation being used while reasoning.

Taking an information processing approach, Vessey (1991) stated that more effective problem solving will occur when the complexity of the task environment is lower. This complexity is reduced when the “problem-solving aids,” like external representations, match well to the cognitive processes required to solve the task. When a task and external representation match, cognitive fit occurs, meaning that the cognitive processes that are activated by the task and the external representation match one another, leading to a consistent mental representation of the task. When a task and external representation do not match, transformations of the mental processes and/or mental representation are necessary, which will lead to a lower efficiency of the problem solving process. Cognitive fit theory was originally designed to explain the inconsistent results concerning the efficiency of solving tasks using graphs compared to tables (Vessey, 1991), but has later been successfully applied to, for instance, the extraction of information from various representational information displays (Zhang, 1996) and software comprehension and modification (Shaft & Vessey, 2006).

Without referring explicitly to Cognitive fit theory, the notion of the existence of cognitive fits is also acknowledged in the mathematics education literature, where it has often been stated that each representation emphasizes a specific aspect of a mathematical concept (e.g., Goldin, 2002). Leavy (2006) made the same claim within the domain of representations for data distributions, arguing that graphical representations facilitate a focus on specific aspects of distributions that one would miss when only looking at descriptive statistics. Descriptive statistics can, for instance, facilitate a focus on individual values, while graphical representations reveal the shape of a distribution better. There is, however, very little known about the way people interpret various aspects of distributions using different representations (Leavy, 2006). We wish to extend Leavy’s (2006) claim, and argue, first, that every representation may facilitate a focus on one or more specific aspects of a distribution that one would miss when studying the distribution in another representation, and, second, that representations may also *detract* attention from other aspects of a distribution. A boxplot, for example, may detract one’s attention from the global shape of the distribution, or the data in the whiskers because of the prominence of the box, whereas at the same time facilitating a focus on the median. Thus, depending on the particular question, a given representation may help or hinder in answering the question.

The design principles of Tufte (1983) and Tversky (1997) outlined above were developed to explain why a given representation, in general, can be easier to interpret than another representation. But given the assumptions of the existence of cognitive fit as described above, they may also be applied at a more subtle level, and help explain why a particular representation can be more helpful in solving a certain task than another. For example, when looking at a boxplot to judge the skewness of a distribution, a viewer may be misled by the area of the box indicating that the density of observations is higher in the smaller area whereas the opposite is the case for histograms. When looking at the median, however, the area of a box will most likely not influence students’ reasoning in a negative way. This way, a boxplot could be a “fit” for tasks concerning the median, but a “misfit” for tasks concerning skewness.

In previous research, no systematic (theoretical or empirical) comparison of different representations for a given set of tasks has been made concerning reasoning about (data) distributions. There is no clear empirical evidence on cognitive fits between tasks and representations in distributional reasoning. As explained above, Cognitive fit theory itself is only a generic framework, and is not specific enough to make clear predictions on which representations will probably work best for certain tasks in a particular cognitive task domain, such as distributional reasoning.

#### **1.4. RESEARCH GOALS**

In contrast to the studies in the above literature overview, the present study focused on the impact of the design of the external representations on students’ distributional reasoning. Instead of using only one representation for each task (as was done in the previous studies), we used the type of representation as a between-subject variable, and ensured each type of task solved with each type of representation.

Our first research question focused on the existence of matches between particular external representations of data distributions on the one hand, and particular aspects of distributions that have to be compared on the other: *Is it possible to explain achievement on distributional reasoning items based on the combination of an item and the given representation and therefore, to identify matches between certain items and certain representations?*

Our second question focused on the nature of incorrect reasoning that students show, which may provide a process-based explanation for the existence or absence of matches as put forward in the first research question: *Which difficulties do students have in interpreting different kinds of representations of data distributions? Are these difficulties specific to certain representations and/or do they relate to certain items?* We expected to find some of the difficulties to be similar to those already discussed in our literature review, but new ones could come forward. Concerning histograms, we predicted that some students would consider the height differences of the bars to compare variation, think that the height of bars represents value instead of the frequency or proportion, think of the horizontal axis as a timeline, ignore the classes in a grouped histogram, and use the frequency of the mode to compare means. Concerning boxplots, we predicted that some students would misinterpret the boxplot by thinking that a larger area represents more observations. Concerning dotplots, we did not expect any specific misinterpretations, although a local view on data might occur more frequently with this representation, as all individual data points are visible, which is not the case in the other representations. Not only for the graphical representations, but also with the descriptive statistics we expected some misinterpretations. More specifically, we predicted that some students would display a local as opposed to a global perspective on distributions, as the descriptive statistics do not allow them to see the distributions in a glance leading them to focus on one or more numbers in the list of descriptive statistics.

## 2. METHOD

### 2.1. PARTICIPANTS

First-year university students of educational sciences ( $n = 167$ ; 158 female, 9 male) from the KU Leuven in Belgium participated in this study, in return for credit on a course assignment. All students in the course participated. The strong overrepresentation of females is due to student enrollment patterns. Prior to their participation, all students had followed the same introductory statistics course that covered descriptive statistics (including several measures of centre and variation), graphical representations (including bar graphs, dotplots, histograms, and boxplots), and distributions (both empirical and theoretical distributions), amongst other topics.

### 2.2. MATERIALS, DESIGN, AND PROCEDURE

Each version of the paper-and-pencil test consisted of eight items in which students were asked to compare the distribution of life spans for two brands of batteries. Comparisons had to be made with respect to an aspect of the two distributions that was specified in the question: the mean, the median, the variation, or the skewness. We are aware of the fact that mean and median are measures that can both be used as measures of centre, while variation and skewness are concepts that can be expressed by numerous measures. This difference is especially relevant for the descriptive statistics, as the mean and the median can be directly read from the list, whereas one needs to know that the standard deviation is a measure of variation and be familiar with the way a skewness measure needs to be interpreted.

All distributions were unimodal. In half of the comparisons, two symmetrical distributions had to be compared (symmetrical items), while in the other half, a symmetrical and an asymmetrical (skewed) distribution had to be compared (asymmetrical items). Each student received two comparison items, one symmetrical comparison and one asymmetrical comparison, for each distribution aspect (mean, median, variation, or skewness).

Four test versions were developed by manipulating the type of graphical representation for the two distributions in an item. Only one kind of external representation was used for the two distributions in an item: either two dotplots, two grouped histograms, two boxplots, or two tables of

descriptive statistics (mean, median, mode, standard deviation, quartiles, minimum, maximum, and skewness). These four representations were chosen because they each have a different level of abstractness (Roth & Bowen, 1994), were known to our participants, and were widely used. Every student encountered each type of representation for two items in an eight-item test version, with each item referencing a different aspect of the distribution (see Table 1).

*Table 1. The design characteristics of the four test versions*

ITEM	Version 1	Version 2	Version 3	Version 4
1	Descriptives	Box plot	Descriptives	Histogram
	Median	Median	Skewness	Mean
	Symmetric	Symmetric	Asymmetric	Symmetric
2	Dotplot	Dotplot	Histogram	Descriptives
	Variation	Skewness	Median	Median
	Asymmetric	Symmetric	Symmetric	Asymmetric
3	Dotplot	Histogram	Descriptives	Dotplot
	Mean	Median	Mean	Skewness
	Symmetric	Asymmetric	Asymmetric	Asymmetric
4	Descriptives	Descriptives	Histogram	Dotplot
	Skewness	Variation	Skewness	Median
	Asymmetric	Asymmetric	Asymmetric	Symmetric
5	Histogram	Histogram	Box plot	Histogram
	Skewness	Variation	Mean	Variance
	Symmetric	Symmetric	Symmetric	Asymmetric
6	Histogram	Box plot	Dotplot	Box plot
	Mean	Skewness	Median	Skewness
	Asymmetric	Asymmetric	Asymmetric	Symmetric
7	Box plot	Dotplot	Box plot	Box plot
	Median	Mean	Variation	Mean
	Asymmetric	Asymmetric	Asymmetric	Asymmetric
8	Box plot	Descriptives	Dotplot	Descriptives
	Variance	Mean	Variation	Variation
	Symmetric	Symmetric	Symmetric	Symmetric

For each item the students were asked to choose the correct answer from five given alternatives, which were always the same: “A,” “B,” “equal,” “you cannot be sure,” and “I don’t know.” In 11 of the 32 items, however, more than one response was considered as correct (i.e., “you cannot be sure”) as it is possible to find datasets with different summary statistics but the same graphical representation as the one presented in the item. However, the frequency with which students gave the “you cannot be sure” response suggests that the students did not use this reasoning often. Students also had to explain how they determined each answer. Four sample items are provided in Figure 1. The representations were created with Fathom™ 2.11 (Finzer, 2005).

The test was administered in groups of approximately 25 students in a classroom. Students were asked to work individually and at their own pace on the eight items. Most students finished the test within 20 minutes.

### 2.3. ANALYSIS

To test for main effects and interaction effects we used generalized linear mixed models (the *glimmix* procedure in SAS™). This enabled us to take the possible association between the different responses of the same participant into account. For the tests of main effects of representation per item we used logistic regression analysis in SAS. In some analyses we used Firth’s method of penalized maximum likelihood estimation to deal with quasi-complete separation of the data points (Firth, 1993).

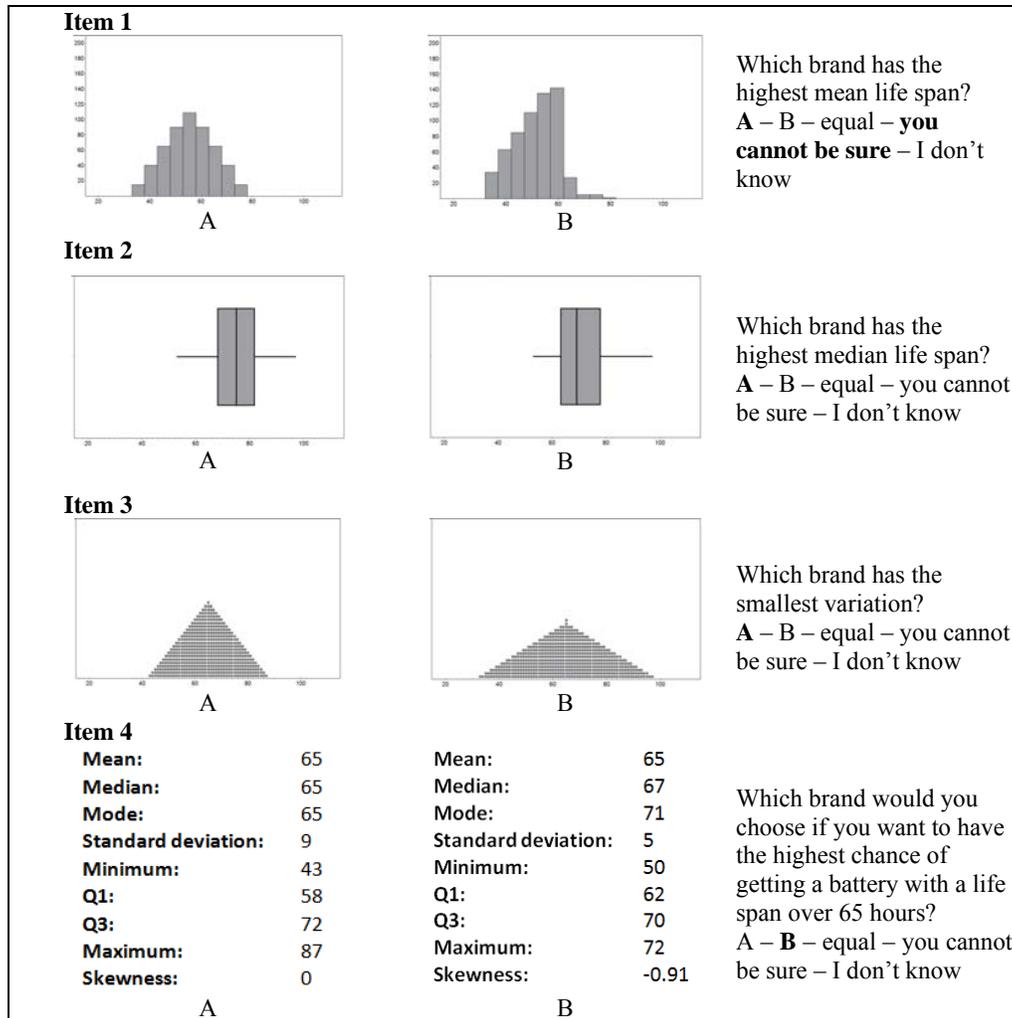


Figure 1. Four sample items. Item 1 compares the means of a symmetrical and an asymmetrical distribution presented as grouped histograms. Item 2 compares the medians of a symmetrical and an asymmetrical distribution presented as boxplots. Item 3 compares variation of two symmetrical distributions presented as dotplots. Item 4 compares the skewness of a symmetrical and an asymmetrical distribution presented as descriptive statistics. Correct responses are marked in bold.

For the analysis of students' explanations we coded all occurring explanations. In this coding process, we looked only at the explanations of the students, and thus neglected the correctness of their answers, which means that a correct response could be accompanied by an incorrect explanation. The coding scheme was developed during the coding process itself, creating a new category for each new error that was observed. The first author coded all 1,326 explanations and the second author coded 320 observations (24%; 10 randomly selected observations per task-representation combination) to test the inter-rater reliability. The agreement between the two raters was 94% and Cohen's kappa was calculated at .97, suggesting a high inter-rater reliability. Discrepancies between the two coders were resolved through discussion. The results of the coding process are discussed in the Results section. In this discussion we limit ourselves to the explanations that occurred in 10% or more of the cases or explanations that were particularly interesting but occurred less frequently.

### 3. RESULTS

Overall, 58% of responses were correct. On average, students answered 4.28 out of eight items correctly, with a standard deviation of 1.96. Table 2 gives an overview of the percentages of correct responses per task and representation. A generalized linear mixed model with correctness of response as the dependent variable showed a main effect of task,  $F(7, 1129) = 21.17, p < 0.001$ , a main effect of representation,  $F(3, 1129) = 11.86, p < 0.001$ , and an interaction effect of task and representation,  $F(21, 1129) = 2.83, p < 0.001$ . This indicates that different representations led to different percentages of correct responses, and that the effect of representation was not the same for every item. Per task, logistic regression analyses (the results of which are summarised in Table 2) showed a main effect of representation for four out of eight tasks.

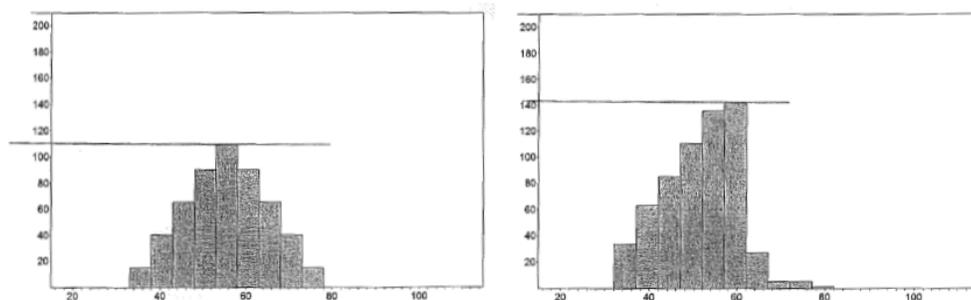
*Table 2. Percentages of correct responses for all items, together with the results of the logistic regression analyses*

	Dotplot	Histogram	Boxplot	Descrip.	$\chi^2$	df	n	p	$R^2_{max}$ <i>rescaled</i>
Mean – sym.	59%	34%	68%	83%	19.93	3	166	< .001	.17
Mean – asym.	45%	38%	71%	83%	21.37	3	165	< .001	.18
Median – sym.	92%	94%	83%	91%	2.83	3	166	.418	.03
Median – asym.	48%	55%	69%	95%	17.01	3	167	< .001	.22
Variation – sym.	72%	59%	74%	74%	3.11	3	164	.376	.03
Variation – asym.	61%	57%	51%	76%	5.47	3	166	.140	.05
Skewness – sym.	18%	34%	38%	23%	5.04	3	167	.169	.04
Skewness – asym.	42%	35%	17%	68%	19.70	3	165	< .001	.18

Further discussion of the results with respect to the aspect of the distribution is presented in the following two subsections.

#### 3.1. COMPARING MEANS

A main effect of representation was found for both mean comparison tasks. The low performance for the histogram representation is notable. Students' explanations helped to explain this low performance. When comparing the means of two symmetric distributions, 23% of the students looked at the frequency of the mode in order to make a decision regarding the means, thinking that the distribution where the mode was observed more frequently had a higher mean. This kind of reasoning is illustrated by a student's reasoning in Figure 2. This can be related to Tversky's design principle of



*Figure 2. A student's explanation for answering that the mean of the right histogram was higher than the mean of the left histogram: "Battery has the highest top. Top = mean."*

direction; these students interpreted the height differences instead of the horizontal spread of the histograms. Furthermore, when using the histogram in the symmetric comparison item, 16% of the students used the variation of the distributions to compare the means, while this did not happen systematically with any of the other representations. There was some variety in the explanations given by the students who considered the variation: Some thought that a distribution with a larger range or larger variation had a higher mean, while others reasoned the opposite way.

### 3.2. COMPARING MEDIANS

The symmetric median comparison tasks yielded relatively high accuracy rates for all representations and main effect of representation was not found. For the asymmetric item, a main effect of representation was found, with the dotplot and histogram representation leading to more incorrect responses. With respect to the asymmetric dotplot item, 12% of the students stated that it was not possible to find the answer using this representation. With respect to the histogram representation, many students compared the height of the modes instead of the medians. This happened in both the symmetric (20%) and asymmetric (29%) items and is the same misinterpretation as displayed in Figure 2, in which students used the height of the modal bars to compare the means instead of the medians. Again, this misinterpretation can be related to Tversky's design principle of direction. Also the asymmetric boxplot item had a lower correct response rate than the descriptive statistics item in which the median could be read off directly, which can be explained by two misinterpretations the students displayed. First, five percent of the students took into account the (irrelevant) vertical dimension in the boxplot, some by saying they could not answer the item because there were no numbers or labels on the vertical axis of the boxplots, and others by saying that the median of the two boxplots was the same because the whiskers and box of both boxplots were drawn to the same height in both boxplots. These misinterpretations can again be related to Tversky's principle of direction: Students tried to take into account the vertical dimension although a boxplot only has relevant information represented in the horizontal dimension. Second, in the symmetric boxplot item, 17% of the students used the variation of the distributions to come to a conclusion about the median, some by reasoning that a larger variation means a larger median, others by reasoning the other way around.

### 3.3. COMPARING VARIATION

With respect to the variation comparison tasks, a main effect of representation on accuracy was not found, but for each of the four representations there was a considerable number of errors. Two major misinterpretations were identified. First, when comparing the variation of two histograms, eight percent of the students comparing two symmetric histograms looked at the height differences of the different bars. They reasoned that larger height differences between the bars lead to a higher variation (see a student's explanation in Figure 3 in which a student illustrated this reasoning in an asymmetric item). This misinterpretation can be linked to Tversky's design principle on the use of direction:

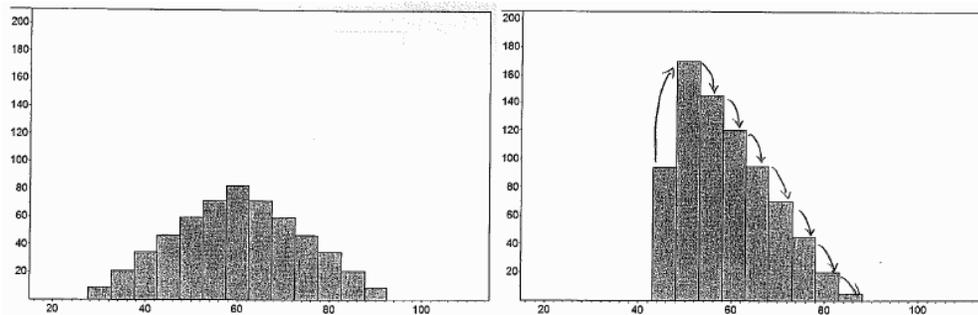


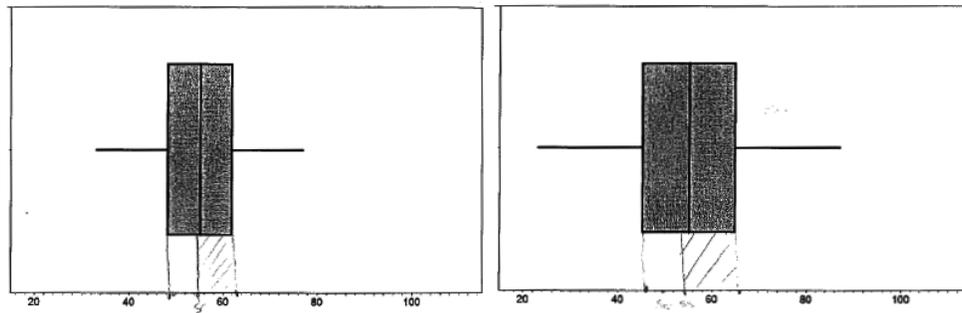
Figure 3. A student's explanation for why the variation in the right histogram is larger than the variation in the left histogram: "The transition from bar to bar (see arrows) is larger in the right histogram. Variation = larger."

Students relied on differences in the vertical dimension instead of looking at the horizontal spread of the histograms.

Second, in the asymmetric boxplot item, 24% of students used the skewness to assess the variation. They all reasoned that a symmetrical distribution would have a smaller variation than an asymmetrical distribution.

### 3.4. COMPARING SKEWNESS

For the skewness comparison tasks, we found a main effect of representation only in the asymmetric task. Overall, accuracy was relatively low. Three misinterpretations related to the graphical features of the representations were found. First, with respect to the boxplot items, 14% and 29% of students in the symmetric and asymmetric item, respectively, used the area of the box to come to a conclusion about the number of observations before and after the median: They reasoned that a larger area (of part) of the box represented more observations than a smaller area (of part) of the box (see Figure 4 for the explanation of a student). This misinterpretation is an instantiation of Tversky's design principle on the use of space: Students tried to interpret space in a natural way by assuming that it represented frequency or proportion of observations. Second, also with respect to comparing boxplots, seven percent of the students ignored the whiskers of the boxplot in the symmetric item. Again, this can be related to Tversky's design principle on the use of space, as the whiskers take much less space than the box, making students think the whiskers represent fewer or no observations. Third, when confronted with histogram or dotplot items, students compared the absolute frequencies on both sides instead of the relative frequencies. Of these students, 15% did this with the symmetric histogram item, and six percent and 33% with the asymmetric dotplot and histogram items respectively. This last misinterpretation can be related to the graphical design of the representations, as in these representations the data are shown more directly (including absolute frequencies) and are less summarized than the other representations, where absolute frequencies are not visible.



*Figure 4. A student's explanation for why there is a larger chance in the right boxplot to have a battery with a life span longer than 65 hours: "In comparison to the left boxplot, you have a bigger chance of a battery with a life span longer than 65 hours (see the area)."*

In addition to the three misinterpretations related to the graphical features of the representations, we found two other misinterpretations when comparing skewness. First, students often used variation in comparing the skewness of two distributions. These students reasoned that the distribution with the largest range and/or the highest maximum would also have the highest chance of getting a battery with a lifespan above the mean. With the symmetric item this occurred for all four representations: dotplot (39%), histogram (30%), boxplot (31%), and descriptive statistics (63%). With the asymmetric item this happened with all representations except the histogram: dotplot (23%), boxplot (32%), and descriptive statistics (30%). Second, students used various measures of centre to compare the skewness of two distributions. These students reasoned that the distribution with the higher mean, median, or mode would also have the highest chance of getting a battery with a lifespan above the mean. Twelve percent of the students did this in the symmetric boxplot item, while five percent gave

this explanation in the asymmetric boxplot item, and 65% with the asymmetric descriptive statistics item.

## 4. DISCUSSION

The present study was set up to investigate the role of representations in distributional problem solving. In the following subsections we summarize the results by linking them to Cognitive fit theory and graph design principles, and discuss the implications for research and educational practice.

### 4.1. COGNITIVE FIT THEORY

Our results show that the accuracy rates for comparing various aspects of two distributions depend on the representation that accompanies the task. Moreover, the effect of representation is not the same for every task: Some tasks are solved better with one representation, whereas others are solved better with other representations. At least for the students in our study, there seems to be a cognitive fit between certain external representations and certain distribution comparison tasks.

However, the percentages of correct responses were relatively low overall. Ideally, in order to support the existence of cognitive fit, accuracy rates should be very high for one or more representations and markedly lower for other representations. When strong claims about matches between representations and tasks could be made based on the data of the current study, however, these matches were not in terms of how well the students solved a certain task using a certain representation, but rather in terms of how poor they were at solving it with other representations. Moreover, even students who gave a correct response did not always base their response on correct reasoning. Basing cognitive matches between tasks and representations on accuracy rates alone is hence rather dangerous.

Therefore, taking only characteristics of the task and representation into account does not seem sufficient to predict which external representation is best to solve a particular task. Acevedo Nistal, Van Dooren, Clarebout, Elen, and Verschaffel (2009) stated that in the domain of representational flexibility, besides characteristics of the representation and of the task, characteristics of the context and the representational fluency of the problem solver should be taken into account when trying to establish matches between items and representations. This last factor means that, although task and representation might fit very well according to the Cognitive fit perspective, the fluency of a person with that specific representation can hinder him or her in solving the task correctly, so that the apparent fit is actually a misfit for that person. Based on the results of this study, straightforward claims cannot be made about the existence of matches between distribution comparison tasks and representations, as these would be based on the many mistakes students made, instead of on their ability to use the representations. With expert users of these representations, representational fluency would be a less confounding variable, possibly making it easier to find concrete matches between tasks and representations.

### 4.2. GRAPH DESIGN PRINCIPLES

The graph design principles of Tufte (1983) and Tversky (1997) discussed earlier were useful to interpret the errors and misinterpretations students displayed. First, as predicted by Tufte's quality indicators (the data-ink ratio and the data-density index), the boxplot representation did not yield very high accuracy rates. However, the histogram appeared to be even worse, which was not predicted by these indicators, as the data-ink ratio and data-density index of histograms are higher than those of boxplots. Second, Tversky's (1997) graph design principle concerning the natural use of space proved to be very helpful to understand students' misinterpretations. Students interpreted a larger area in a boxplot to represent more observations, and ignored the whiskers of boxplots, thinking that these small elements did not represent (m)any observations. Concerning dotplots and histograms, students employed a local view on the distribution more often than with the other representations, as predicted. Third, many of the students' incorrect explanations were due to a focus on the vertical dimension of the graphs instead of on the horizontal dimension, which is in accordance with Tversky's (1997) second design principle. This misinterpretation was displayed by students' comparison of the height

differences of the bars in a histogram to compare variation, and their comparison of the height of the modal bars in histograms to compare means and medians.

Other misinterpretations could not be explained directly by the graph design principles of Tufte (1983) and Tversky (1997), but did seem to be caused by specific characteristics of the representations. Mainly, we observed that many students used characteristics of a distribution that were irrelevant to the task. The characteristic that was utilized incorrectly differed to a large extent per task and per representation and was, apparently, often based on superficial cues of the representation that led students' attention towards that specific feature. So, with respect to Leavy's (2006) suggestion that different representations can facilitate the focus on specific aspects of distributions, the results show that this focus may also *hinder* students from arriving at the correct answer: The cues provided by a certain representation may lead to a focus on an irrelevant characteristic and thus to an incorrect solution.

### 4.3. IMPLICATIONS FOR RESEARCH AND EDUCATIONAL PRACTICE

**Research** The results have several implications for future research. First, they show the important role of external representations upon distributional reasoning: When a person does not completely understand a given representation, it is likely that he or she will not be able to use it properly in solving distribution tasks, and with certain representations, the chances of solving a certain task correctly are considerably higher than with other representations. Moreover, every combination of a certain representation and a certain task can lead to very specific misinterpretations. Many previous studies of distributional reasoning, however, only used one representation in their items (e.g., Meletiou & Lee, 2002; Pfannkuch, 2006; Prodromou & Pratt, 2006; Watson & Moritz, 1999). We think that not taking into account the variety of representations that are provided to students when studying their distributional reasoning abilities is problematic: Conclusions about students' distributional reasoning are in that case most likely not generalizable to problems involving other representations of distributions. One could also wonder whether those studies observed distributional reasoning, or whether they alternatively observed how well students understand external representations for distributions. One might even question whether it is possible to observe distributional reasoning *directly*, as it is difficult to make a distinction between difficulties that occur because of a misinterpretation of the representation and difficulties that occur due to a misunderstanding of concepts related to statistical distributions. This problem occurs in research on mathematical problem solving as well, but so far has not been addressed very often in the statistics education literature. A research design in which representations are systematically varied across participants, like the one presented here, can circumvent this problem.

Second, our results suggest that, when looking for matches between tasks and representations, other factors, such as representational fluency of students who solve the tasks, should also be taken into account. In this case, the representational fluency of the students appeared to be very low, compromising the matching of representations to tasks based on our data.

Third, consideration of the study's limitations may yield guidelines for further research. Individual differences between students were not examined. It would be interesting to study how subject characteristics, such as representational flexibility and mathematical ability, affect the match between representation and task. Furthermore, the low representational fluency of the participants impeded the matching of representations and tasks. Undertaking a similar study with expert participants (i.e., statisticians) may help to identify better matches between tasks and representations because fluency would play a smaller role. Also, the participants were from a very specific group (educational sciences) and were mainly female. Similar studies with different student populations, for example from other fields of study with less gender imbalance, could be interesting. Finally, process data were not collected, only self-reported reasoning. Future research could use individual interviews to get a better view of students' reasoning and whether incorrect reasoning is the result of superficial, mindless reasoning, or of deeper-rooted misconceptions.

Fourth, another interesting line of future research lies in the design of boxplots. Tufte (1983) proposed several redesigns of the classic boxplot in which the interquartile range is no longer presented two-dimensionally but for example as a thin line or an open space between the whiskers, with a bullet indicating the position of the median. Based on the design principles of Tufte (1983) and

Tversky (1997) and our analysis of students' explanations, we expect that these redesigns would probably indeed help students to interpret boxplots correctly.

**Educational practice** A general conclusion from the study results is that the use of the four types of representation by students is not without problems and should be reassessed. The results also hold several specific implications for educational practice. First, we have argued that representational fluency is very important: Many misinterpretations of the representations in the study led to incorrect solutions being given to the items. When students are not able to interpret representations of data distributions properly, it is unlikely that they will be able to understand more complex underlying concepts, such as sampling and theoretical distributions. More instructional attention should hence be paid to the development of fluency with these representations.

Second, when we know what difficulties students have when learning to use certain representations, we can explicitly address these issues in order to try to diagnose, prevent, and remediate them. A possible strategy for solving these problems could be the use of multiple representations of the same distribution. This would allow students to confront information they extract from a certain representation with information from other representations, and learn to translate and switch between representations.

### ACKNOWLEDGEMENTS

Stephanie Lem holds a PhD fellowship of the Research Foundation – Flanders (FWO). This research was partially supported by the grant G063709N “Representational adaptivity in mathematical thinking and learning: analysis and improvement” from the Fund for Scientific Research Flanders.

### REFERENCES

- Acevedo Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: A critical review. *ZDM Mathematics Education*, 41(1), 627–636. doi:10.1007/s11858-009-0189-1
- Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (1998). Analysing the costs and benefits of multi-representational learning environments. In M. W. van Someren, P. Reimann, H. P. A. Boshuizen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 120–134). Amsterdam: Pergamon.
- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (April, 2002). *The resilience of overgeneralization of knowledge about data representations*. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA.
- Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable* (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.
- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63.  
[Online: [http://iase-web.org/documents/SERJ/SERJ3%282%29\\_BenZvi.pdf](http://iase-web.org/documents/SERJ/SERJ3%282%29_BenZvi.pdf)]
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1/3), 35–65.  
doi: 10.1023/A:1013809201228
- Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, 23(3), 369–381.
- Cohen, S., Smith, G., Chechile, R. A., Burns, G., & Tsai, F. (1996). Identifying impediments to learning probability and statistics from an assessment of instructional software. *Journal of Educational and Behavioral Statistics*, 21(1), 35–54.

- Cooper, L. L., & Shore, F. S. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education*, 16(2). [Online: <http://www.amstat.org/publications/jse/v16n2/cooper.html>]
- De Vries, E., Demetriadis, S., & Ainsworth, S. (2009). External representations for learning: Headed towards a digital culture. In N. Balacheff, S. Ludvigsen, T. De Jong, A. Lazonder, & S. Barnes (Eds.), *External representations for learning* (pp. 137–153). Dordrecht, The Netherlands: Springer.
- delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy*, Auckland, 2-7 July 2005, [CD-ROM, with video segments]. Brisbane, Australia: University of Queensland.
- Finzer, W. (2005). *Fathom<sup>TM</sup>: Dynamic Data* (Version 2.11) [software]. Emeryville, CA: KCP Technologies.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, 119(1), 63–76.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397–409). Dordrecht, The Netherlands: Kluwer. doi: 10.1007/1-4020-2278-6\_17
- Garfield, J. B., & Ben-Zvi, D. (2008). Learning to reason about distribution. In J. B. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and teaching practice* (pp. 165–186). Dordrecht, The Netherlands: Springer. doi: 10.1007/978-1-4020-8383-9\_8
- Goldin, G. A. (2002). Representation in mathematical learning and problem solving. In L. D. English (Ed.), *Handbook of international research in mathematics education* (pp. 197–218). Mahwah, NJ: Erlbaum.
- Kaput, J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 515–556). New York: Macmillan.
- Kelly, A. E., Sloane, F., & Whittaker, A. (1997). Simple approaches to assessing underlying understanding of statistical concepts. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 85–90). Amsterdam: IOS Press.
- 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.
- Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (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: Proceedings of the 1996 IASE Round Table Conference* (pp. 151–168). Voorburg, The Netherlands: International Statistical Institute.
- Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). *Data seen through different lenses*. Unpublished manuscript.
- Kosslyn, S. M. (2006). *Graph design for the eye and mind*. New York: Oxford University Press, Inc.
- Leavy, A. M. (2004). Indexing distributions of data: Preservice teachers' notions of representativeness. *School Science and Mathematics*, 104(3), 119–134.
- Leavy, A. (2006). Using data comparisons to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal*, 5(2), 89–114. [Online: [http://iase-web.org/documents/SERJ/SERJ5%282%29\\_Leavy.pdf](http://iase-web.org/documents/SERJ/SERJ5%282%29_Leavy.pdf)]
- Meletiou, M., & Lee, C. (2002). Student understanding of histograms: A stumbling stone to the development of intuitions about variation. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
- Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
- Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27–45.

- [Online: [http://iase-web.org/documents/SERJ/SERJ5%282%29\\_Pfannkuch.pdf](http://iase-web.org/documents/SERJ/SERJ5%282%29_Pfannkuch.pdf)]
- Prodromou, T., & Pratt, D. (2006). The role of causality in co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, 5(2), 69–88.  
 [Online: [http://iase-web.org/documents/SERJ/SERJ5%282%29\\_Prod\\_Pratt.pdf](http://iase-web.org/documents/SERJ/SERJ5%282%29_Prod_Pratt.pdf)]
- Robbins, N. B. (2005). *Creating more effective graphs*. Hoboken: Wiley.
- Roth, W., & Bowen, G. M. (1994). Mathematization of experience in a grade 8 open-inquiry environment: An introduction to the representational practices of science. *Journal of Research in Science Teaching*, 31(3), 293–318. doi: 10.1002/tea.3660310308
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–70). Cambridge: Cambridge University Press.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction*, 13(2), 141–156. doi: 10.1016/S0959-4752(02)00017-8
- Shaft, T. M., & Vessey, I. (2006). The role of cognitive fit in the relationship between software comprehension and modification. *MIS Quarterly*, 30(1), 29–55.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire: Graphic Press.
- Tversky, B. (1997). Cognitive principles of graphic displays. In M. Anderson, (Ed.), *Reasoning with diagrammatic representations II: Papers from the AAAI 1997 Fall Symposium* (pp. 116–124). Technical Report FS-97-03. Menlo Park, CA: AAAI Press  
 [Online: <http://www.aaai.org/Papers/Symposia/Fall/1997/FS-97-03/FS97-03-015.pdf>]
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2), 219–240. doi:10.1111/j.1540-5915.1991.tb00344.x
- Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168. doi: 10.1023/A:1003594832397
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.  
 [Online: [http://iase-web.org/documents/SERJ/SERJ5%282%29\\_Wild.pdf](http://iase-web.org/documents/SERJ/SERJ5%282%29_Wild.pdf)]
- Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human Computer Studies*, 45(1), 59–74.

STEPHANIE LEM  
 Dekenstraat 2 P.O. Box 3773  
 3000 Leuven  
 Belgium