Int J of Sci and Math Educ (2015) 13:909–926
DOI 10.1007/s10763-014-9604-3

Combining Multiple External Representations and Refutational Text: An Intervention on Learning to Interpret Box Plots

Stephanie Lem · Goya Kempen · Eva Ceulemans · Patrick Onghena · Lieven Verschaffel · Wim Van Dooren

Received: 27 September 2013 / Accepted: 23 November 2014 / Published online: 19 December 2014
© Ministry of Science and Technology, Taiwan 2014

Abstract Box plots are frequently misinterpreted and educational attempts to correct these misinterpretations have not been successful. In this study, we used two instructional techniques that seemed powerful to change the misinterpretation of the area of the box in box plots, both separately and in combination, leading to three experimental conditions, next to a control condition. First, we used multiple external representations: Histograms were used as an overlay on box plots in order to give students a better insight in the way box plots represent data distributions. Second, we used refutational text to explicitly name and invalidate the area misinterpretation of box plots. Third, we combined multiple external representations and refutational text. A box plot test showed that students in the refutation and combination condition scored statistically significant better than students in the control condition with respect to the misinterpretation of interest. The condition with multiple external representations scored in between. The implications of these results for theory and educational practice are discussed.

Keywords Intervention · Multiple external representations · Refutational text · Box plots

Box plots are often used to graphically depict the distribution of one or more quantitative variables. Recent studies have shown, however, that they are not as easy to interpret as

Electronic supplementary material The online version of this article (doi:10.1007/s10763-014-9604-3) contains supplementary material, which is available to authorized users.

S. Lem (✉) · G. Kempen · L. Verschaffel · W. Van Dooren
Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, PO Box 3773, 3000 Leuven, Belgium
e-mail: stephanie.lem@ppw.kuleuven.be

E. Ceulemans · P. Onghena
Methodology of Educational Sciences Research Group, KU Leuven, Tiensestraat 102, PO Box 3762, 3000 Leuven, Belgium

often thought (e.g. Bakker, Biehler & Konold, 2005; Lem, Onghena, Verschaffel & Van Dooren, 2012, 2013a, b; Pierce & Chick, 2013). Students have more difficulties interpreting box plots than other external representations[1] of data distributions, such as histograms or descriptive statistics (Lem et al., 2013a). Common misinterpretations of the box plot include, for example, thinking that the median line actually represents the mean or ignoring the whiskers in the assumption that no data are represented in this part of the box plot (e.g. Lem et al., 2012, 2013a). In this article, we focus on one specific misinterpretation, namely, the area misinterpretation, which was first identified by Bakker et al. (2005) and later also studied by Lem et al. (2013b) and Pierce & Chick (2013). Students who hold this misinterpretation think that a larger area in a box plot represents a larger proportion of data than a smaller area, while a larger area actually represents the same amount of data but spread out over a larger interval, thus representing a lower density.

In previous studies, several attempts have been undertaken to reduce this misinterpretation, using warnings (Lem et al., 2013b), a short intervention in which box plots were reinvented by students using a bottom-up approach (Lem et al., 2013b), and the alternative designs of box plots proposed by Tufte (1983; Craeghs, 2012). In all cases, the effect on reducing the misinterpretations was at best very limited. In this article, we report on a study in which we tested whether multiple external representations (MERs), refutational text (in which the area misinterpretation is explicitly disproved and a correct interpretation is provided), or the combination thereof improves students' interpretation of box plots. In what follows, we first discuss the area misinterpretation and the available research concerning its origin. Next, we introduce the two teaching techniques that we used in the current intervention: MERs and refutational text. After this introduction, we describe the goals, method, and results of the intervention study that we have conducted.

The Misinterpretation of Box Plots

As mentioned in the introduction, many misinterpretations of box plots have been documented in the research literature (e.g. Bakker et al., 2005; Lem et al., 2012, 2013a, b). In this article, we focus on one specific misinterpretation, called the "area misinterpretation" (Lem et al., 2013b), as this is a frequently occurring misinterpretation that has been studied thoroughly and that has been proven to be very persistent. In order to fully understand this misinterpretation, we first explain briefly how box plots work. Figure 1 shows an example of a box plot. A box plot represents the distribution of a set of observations of one quantitative variable by means of five summary numbers: the minimum, the first quartile (Q1), the median, the third quartile (Q3), and the maximum. This means that a box plot divides the data into four parts that each represents (approximately) the same number of observations.[2] In Fig. 1, the minimum value is 1 (represented by the end of the left horizontal line) and the value of Q1 is 5 (represented by the most left vertical line), meaning that about 25 % of the observed data is situated between 1 and 5. The median of the box plot in Fig. 1 is 9 (represented

[1] 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).

[2] Depending on the exact distribution of the data, not all four intervals will contain exactly 25 % of the data.

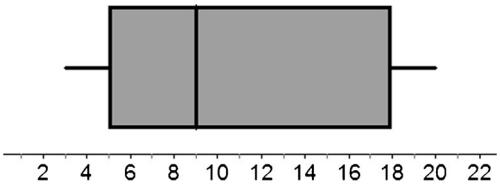


Fig. 1 Example of a box plot

by the middle vertical line), while Q3 is 18 (represented by the most right vertical line). This means that about 25 % of the data is situated between 9 and 18. Despite the interval between 9 and 18 being much larger than the interval between 3 and 5, both intervals contain approximately the same number of observations. This discrepancy between the size of an interval and the number of observations it contains is reported to be difficult not only for students but also for expert users of box plots (Lem, Onghena, Verschaffel & Van Dooren, 2013c). As both Bakker et al. (2005) and Lem et al. (2013a, b, c) discussed, this misinterpretation can be understood by the way we interpret certain salient visual cues. In this specific case, interpreting a larger area as representing more observations is a rather logical interpretation as this is correct in many other circumstances, but not for the box plot.

To study this misinterpretation in depth, Lem et al. (2013b) used a dual-processing perspective (e.g. Evans & Over, 1996; Kahneman, 2000), which has clear similarities with theoretical frameworks that are used in science and mathematics education research such as the intuitive rules theory (e.g. Babai, Brecher, Stavy & Tirosh, 2006; Stavy & Tirosh, 2000). According to the dual-processing perspective, people use two different reasoning processes when confronted with a task: heuristic (or intuitive) processes and analytic processes. While heuristic processes are fast, automatic, and effortless, analytic processes are relatively slow, conscious, and require more working memory capacity. Lem et al. (2013b) tested whether the area misinterpretation could be characterized as fast but incorrect heuristic reasoning by using two types of items for which accuracy and reaction time were recorded. Congruent items are items in which both the expected heuristic reasoning (i.e. the larger area represents more observations) and the correct analytic reasoning (i.e. 25 % of the data is represented in each of the four parts of the box plot) lead to the same (correct) response. This means that many correct responses can be expected and that these responses will be relatively fast. Incongruent items, on the other hand, are items in which the heuristic and the analytic response are not the same, meaning that fewer correct responses are expected and that when a correct response is given, this will take relatively long because more time-consuming analytic reasoning is necessary. This method is very similar to the one used in research on the intuitive rules theory (e.g. Babai et al., 2006). The results of Lem et al. (2013b) show that we can indeed call the area misinterpretation an "area heuristic," both with respect to accuracy and reaction time. This means that when confronted with a box plot, learners automatically focus on the very salient area of the box. When a part of this box is larger than the other parts, they are automatically inclined to think that this means that the largest part represents more data than the other parts. This heuristic reasoning seems to be very strong. More specifically, Lem et al. (2013b) showed that extensively warning students about the possible misleading features of box plots did not improve their interpretation, while

Lem et al. (2013c) showed that even expert users of box plots were negatively influenced by the area heuristic.

Previous Studies in the Field

Various attempts have been undertaken to improve students' interpretation of box plots. Lem et al. (2013b), for example, studied the reasoning mechanisms behind the area misinterpretation and tried to make students cautious of the fact that graphs can be misleading and that they should be aware of this when interpreting box plots, taking their time to carefully reflect on their response. Unfortunately, this did not lead students to a better interpretation of box plots: Students who received this warning performed equally badly as students who were asked to respond within 5 s and did not receive any warning. Second, Craeghs (2012) taught students using the different alternative designs of box plots as proposed by Tufte (1983). In these alternative designs, the area of the box is replaced by lines. The alternative designs did not lead students to a better interpretation of the plot, most probably because the lengths of the different lines were still interpreted incorrectly (i.e. "longer line represents more data," instead of "larger area more data"). Third, Bakker (2004) conducted a series of design experiments using Minitool 2, introduced by Cobb, Gravemeijer, Bowers & Doorman (1997; see also, e.g. Cobb, McClain & Gravemeijer, 2003). This tool starts from a dot plot and overlays a box plot onto it step by step. This allowed the secondary school students that participated in the experiments to actually see that the same number of data points is represented in each of the four parts. Bakker et al. (2005) suggest that the minitool helps students to overcome the area misinterpretation: "Very few students in Bakker (2004) study struggled with this density feature, which indicates that these difficulties might be addressed to some extent by using software that allows displaying dot plots and box plots at the same time" (Bakker et al., p. 171). However, the experiments were rather small scale and the interventions took multiple class sessions, making it difficult to implement them in the regular statistics curriculum. A more recent study (Pierce, Chick & Wander, 2014) aimed at the improvement of teachers' interpretation of assessment reports of schools, in which box plots are often used. A teaching method similar to the minitool was used: The raw data distribution, depicted as drawings of students, was presented below a box plot representing the same distribution. Despite the positive effects this intervention had on some teachers, part of the teachers still showed signs of the area misinterpretation.

In what follows, we will discuss two instructional techniques, namely, multiple external representations and refutational text, that we used in the present study to improve students' interpretation of box plots.

Multiple External Representations

The last few decades, the use of "multiple external representations" (or MERs) is frequently discussed both in the general educational research literature (e.g. Ainsworth, 1999) and in mathematics and science education (e.g. Acevedo Nistal, Van Dooren, Clarebout, Elen & Verschaffel, 2009; Corradi, Elen & Clarebout, 2012). External

representations can take many forms, like written text, graphs, and animations, not only on paper but also, for example, on computer screens. It has been posited that the use of MERs in a learning environment not only improves the learner's motivation but can also support the learner's cognitive processes (Ainsworth, 1999). Three main functions of the use of MERs in learning environments can be distinguished (Ainsworth, 1999): complementing information, constructing deeper understanding, and constraining interpretation. The complementary function of the use of MERs entails that one external representation supplies extra information in a learning environment that is lacking in another external representation. The construction of deeper understanding means that by using MERs, different elements of the learning material are more easily connected, providing an abstraction of the learning material. The third and final function of using MERs, constraining interpretation, is the main function that will be used in our intervention to improve the interpretation of box plots. The latter function implies that the misinterpretation of one representation is prevented by the use of another representation. For example, a horizontal line in a velocity-time graph is often interpreted as "no motion," while it actually represents "uniform motion." By showing a moving object together with its velocity-time graph, this misinterpretation can be prevented (Ainsworth, 2008). This can also be applied to the use of histograms together with box plots; by overlaying box plots onto histograms, it is clearly visible that the same number of data points is represented in each of the four parts of a box plot and that a wider part implies that the data are less densely distributed in that part. The use of the box plot in combination with the histogram can hence be expected to show students that the interpretation of area of the box as a representation of number of observations is incorrect.

Empirical evidence that MERs work is available for different content domains, like solving geometry problems (Van Dooren, De Bock, Hessels, Janssens & Verschaffel, 2004), chemistry (Corradi et al., 2012; Waldrip & Prain, 2012), and genetic reasoning (Tsui & Treagust, 2003). Despite the potential benefits of MERs, various drawbacks have also been reported (Acevedo Nistal et al., 2009). It is, for example, important that the learner is able to interpret all representations in the learning environment correctly. Also, interpreting several representations at the same time and coordinating or switching between them asks a lot of a learner's working memory capacity and attention. Unfortunately, histograms are not without interpretation difficulties either (e.g. Lem, Onghena, Verschaffel & Van Dooren, 2014). It is unlikely, however, that the difficulties observed in previous research negatively influence the interpretation of the histogram—box plot combination in the present study; Lem et al., 2014 showed that while students show specific misinterpretations of the histogram, they are very well able to read the frequency of observed values from the histogram. This is the most important ability needed for the interpretation of histograms in the present study. On top of that, histograms were also explained in the intervention, trying to reduce the possibility of misinterpretations concerning the histograms.

Refutational Text

A refutational text is a text that explicitly introduces a common (but incorrect) conception, then refutes it and offers an alternative (correct) theory (Hynd, 2001). An

example of a refutational text from Palmer (2003), used to teach children that all creatures have a role in the ecological system, reads as follows:

> […] some people think that living things do not have a role if they don't seem to do much, like starfish. But I think starfish do have a role because they probably eat little sea animals. […] All creatures eat things, so they all have roles. (p. 670)

This type of texts has been used mainly in the conceptual change research tradition (Sinatra & Broughton, 2011). This instructional method is used frequently in research on the remediation of misinterpretations and misconceptions in science and mathematics education (Tippett, 2010). Studies on the effect of refutational text have been conducted in different areas. McCrudden & Kendeou (2012), for example, showed how refutational text revised students' incorrect conceptions about Newton's first law, while Gill, Ashton & Algina (2004) reported the positive effects on preservice teachers' epistemological beliefs about teaching and learning in mathematics. Despite these promising results, refutational text is still rarely used in science and mathematics textbooks (Tippett, 2010).

In order to explain why refutational text works, Van Den Broek & Kendeou (2008) propose that co-activation of conceptions plays a crucial role. More specifically, refutational text makes it possible for readers to simultaneously activate incorrect prior knowledge and correct conceptions in order to help them uncover the inconsistency between both. This co-activation is seen as a crucial step in the possible improvement of incorrect conceptions, as it helps to detect them and at the same time provides a correct conception that can replace the incorrect one(s).

Refutational text consists of two essential elements to which one extra element can be added (Tippett, 2010). A first is the explicit naming of the incorrect interpretation that is being addressed. This is very uncommon in most traditional textbooks. A second element is that the refutational text states that this interpretation is incorrect and why and that it provides a new, correct interpretation. A third, but non-essential, element that can be added to refutational text is a refutational cue. This cue alerts the reader that a refutation of the incorrect interpretation will follow, for example, by saying, "this is incorrect."

Goal of the Present Study

The goal of the present study is to find a new way to improve students' interpretation of box plots, given that the three studies discussed above were either not successful or used a very time-intensive intervention. We test the effect of two instructional techniques: MERs and refutational texts. The idea to introduce MERs originates from Bakker's (2004) study with the minitool intervention. In the present study, we work further on the finding that the use of dot plots to teach box plots helped students to some extent. We will also link the box plot to another data representation, but we now work with histograms as the second representation as this fitted better within the curriculum of the students. We acknowledge that using dot plots for the overlay emphasizes each individual data point much better than the bars of histograms, but by using only ungrouped histograms, we feel that this disadvantage of histograms is

diminished, especially because the students in our sample are much more familiar with histograms than with dot plots. With respect to refutational text, we used warnings like Lem et al. (2013b) did, but we extended the warning, made it much more specific to box plots, and used the refutational text format as described by Tippett (2010).

Method

Participants

Participants of the study were 188 first year students of educational sciences (108 females, 7 males) and speech pathology and audiology (71 females, 2 males) at KU Leuven, Belgium.[3] All students were enrolled in the same introductory statistics course and were taught about descriptive statistics prior to their participation. Box plots were not discussed in class until after the intervention and posttest had taken place. Of all students, 82.0 % were already introduced to box plots in secondary education, according to their self-report on their previous box plot experience after the intervention. Participation was voluntary and additional to the regular statistics education program. The statistics teacher (the third author) encouraged the students to take part in the study, using the motivation that the material covered in the study would be an important asset for understanding the materials included in the regular statistics course.

As the sample used was a convenience sample, i.e. all students registered to the introductory statistics course were included in the study, a power analysis was conducted to determine the established statistical power. A power analysis with a significance level of 0.05 using G*Power (Faul & Erdfelder, 1992) revealed that with a sample size of 188 the statistical power for this study was 0.95 for detecting a small effect (i.e. $f=0.25$), whereas the power was ≥0.99 for the detection of a moderate (i.e. $f=0.40$) to large (i.e. $f=0.60$) effect size. We must however note that the statistical power we actually achieved was somewhat lower as our analyses were using a multilevel model while the power analysis necessarily relied on repeated measures ANOVA.

Materials and Procedure

Intervention. The intervention was provided in the form of a home study text that was offered to the participants before the teaching of box plots and histograms would take place in class. Four versions of the text were constructed and provided in four conditions: a control condition, a MERs condition, a refutation condition, and a combination condition. The texts that were used in the three experimental conditions were based on the text of the control condition, except for a few experimental manipulations. In the *control condition* ($n=45$), the text explained histograms and box plots in a standard way that was based on the textbook used in the introductory statistics course of the participants. The area misinterpretation was not named explicitly and both representations were not linked to each other (see Fig. 2 and Appendix A). In the *MERs condition* ($n=43$), histograms were explained first and were then used to explain box plots by overlaying box plots onto histograms. This way, students were able to actually see the amount of data represented in

[3] The gender imbalance is due to student enrollment patterns.

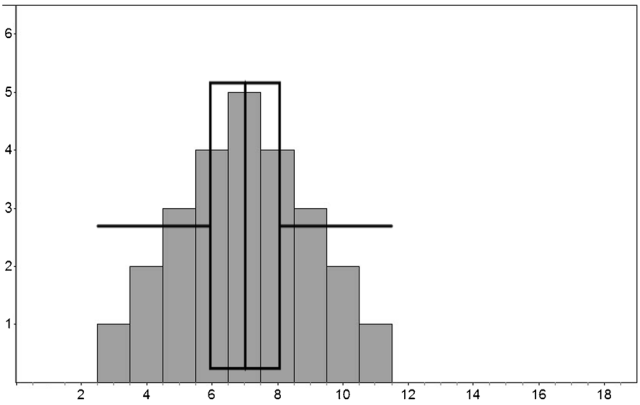


Fig. 2 Example of the box plot overlay on top of a histogram, as used in the MERs condition

each of the four parts of the box plot (see Fig. 3 and Appendix B). This can be categorized as the constraining of interpretation function, as described by Ainsworth (1999). In the *refutation condition* (n=41), refutational texts were added in two text boxes. In line with the definition of refutational text (Tippett, 2010), these two text boxes stated the common incorrect interpretation of box plots, followed by an explicit refutation of this interpretation combined with a correct one. In addition, we added a cue to alert the reader of the misinterpretation: In each of the two refutational text boxes, we did this by adding a small cue sentence in between the misinterpretation and the correct interpretation: "This is incorrect." On top of that, we started each text box with "Warning!!" in order to attract students' attention even more and in the beginning of the instructional text we warned students about the fact that box plots are often misinterpreted by students (see Fig. 3 and Appendix C). In the *combination condition* (n=59), finally, box plots were explained using the same overlay on histograms as in the MERs condition. In addition, the text boxes with warnings from the refutation condition were added (see Fig. 4 and Appendix D).

In all four conditions, the same four exercises were included in addition to the instructional text. The exercises were included in order not only to engage students more in the learning process but also to verify whether students indeed studied the material and whether the possible misinterpretation of histograms would influence the learning about box plots. In the control and refutation condition, the histogram exercises were given directly after the explanation of histograms and the box plot exercises after the explanation of the box plot. In the MERs and combination condition, however,

WARNING!! When you look at a box plot you might think that a larger part of the box represents more results than a smaller part of the box . This is incorrect! In each of the four parts of a box plot (approximately) the same number of observations is represented. The size of the parts of the box do hence not give an indication of the number of represented observations.

A larger part of the box, like in our example between 9 and 12, means that the observations here are more spread out. Between 12 and 14 the observations are hence less spread out. In each of the four parts of this box plot approximately 25% of the exam results are represented.

Fig. 3 Example of refutational text, as used in the refutation condition (translated to English)

WARNING!! You might think that (the box of) the left box plot represents less observations than (the box of) the right box plot as the right box plot has a larger box. This is, like explained earlier, not the case! You cannot read the number of observations off from a box plot. The size of the box only tells you something about the spread of the observations: a larger part of the box shows that the observations are more spread out.

The figures below make the difference between both distributions more clear. You can see the box plots shown above but now with the accompanying histograms.

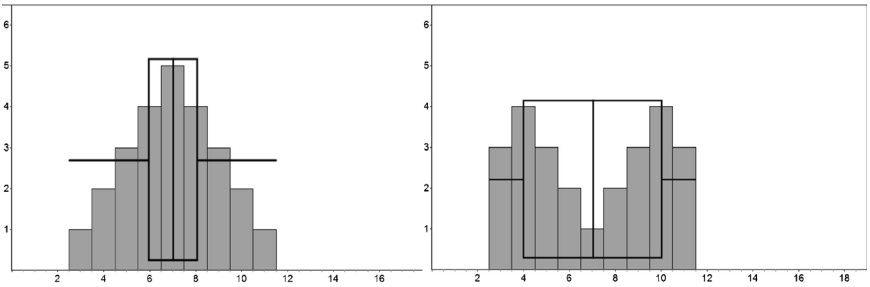


Fig. 4 Example of how refutational text and MERs were combined in the combination condition (translated to English)

all four exercises were provided at the end of the text, as histograms and box plots were explained in an intertwined way. See Appendix E for the exercises. In a first exercise, students were asked to give the minimum, maximum, mode, and direction of skewness for a given histogram. Second, students were asked to draw a histogram of a small given dataset. Third, they were asked to read off minimum, maximum, median, and direction of skewness in a given box plot. Finally, students were asked to draw a box plot of a small given dataset.

Box Plot Test. Differences in learning between conditions were measured by a paper-and-pencil test. Two types of items were included in the test. First, there were four questions concerning students' knowledge on box plots. Students were asked to name the five elements of a given box plot (minimum, Q1, median, Q3, maximum); they were asked to define Q1, to state what percentage of data is presented after Q3, and to say which percentage of the data was represented in a given part of a given box plot. These questions were solely included in the test to make sure each of the four instructional texts was equally proficient in teaching students the theory of box plots. We did not find a main effect of condition on accuracy for the box plot knowledge items, χ^2 (3,564)=0.70, p= 0.552. Box plot knowledge was relatively high after the intervention, with an average of 91.2 %. The most difficult item (in which students had to define Q1) was answered correctly by 76.1 % of the students, while the other three items were answered correctly by 94.2 % (in this item, all elements of the box plot had to be named), 96.3 % (in this item, students had to say what percentage of the data is situated after Q3), and 98.4 % (in this item, students had to say which percentage of the data was highlighted in a given box plot) of the students.

Second, three items directly addressing the area misinterpretation were included. These items were not solvable by the mere reproduction of their knowledge; students had to apply their box plot knowledge to find the correct solution. In these items,

students were asked to provide an explanation for their response. The items came from previous studies on the incorrect interpretation of the area in box plots (Lem et al., 2012). These items will be presented and explained with their respective results in the "Results." Besides the knowledge and area items, some questions concerning personal details, like age and previous box plot knowledge, were included.

Procedure. The four versions of the instructional text were randomly distributed to the students at the end of a statistics class. The students were asked to thoroughly study the text and solve the included exercises at home. They were told that the text covered an important topic of the statistics curriculum and that studying them would help them to understand the contents dealt with in the next class. Using the online learning platform, students who did not attend this specific statistics class could request their text through e-mail. A randomly selected version of the instructional text was e-mailed to the 27 students who requested one. Students were not aware of the existence of different versions of the instructional text.

In the statistics class 1 week after the distribution of the texts, the texts were collected and students received an unexpected test to be completed individually. After completion, we explained the goal and method of the study. Finally, box plots were explained in detail to the students as part of the regular curriculum. This way, we ensured that all students received at least the same instruction.

Analysis

We used logistic regression to study the effect of condition on accuracy for each item separately. To test the effect of condition on all items together, a generalized linear mixed model was used to control for the possible correlation between the different responses of the same student. The explanations students provided for three of the six area misinterpretation items were coded in an inductive way. The coding scheme was constructed during the coding process, adding a new category for each new misinterpretation that was encountered. After the second author scored all 564 explanations according to the coding scheme, the first author coded 150 explanations (26.6 %, 50 randomly selected explanations per item). Interrater reliability was 0.89 (Cohen's kappa), which can be considered to be high (Landis & Koch, 1977). Discrepancies between the two coders were resolved through discussion.

Predictions

We expected a difference between the conditions concerning the accuracy of the area items. More specifically, we anticipated the highest accuracy in the combination condition, as this was the condition in which both MERs and refutational text are used together, and lowest accuracy in the control condition. We expected the refutation condition and the MERs condition to be between the other two conditions. Depending on the specific tasks, differences between the conditions were also expected. More specifically, we anticipated that area item 2 would be solved best by the MERs condition, as in this item, a histogram had to be linked to the corresponding box plot (see Fig. 7).

Results

Three items were used to test whether students interpreted the area of the box plot correctly. A generalized linear mixed model was fitted to all three area items, showing a main effect of condition on accuracy, χ^2 (3,3368)=3.63, p= 0.013, OR=0.43–0.99. Table 1 shows the percentages of correct responses per item and condition. We see that the control condition in total scored worst (54.8 % correct), while the combination condition (73.5 %) and refutation condition (73.2 %) scored best and the MERs condition (67.4 %) scored in between. Dunnett's tests showed that the refutation as well as the combination condition differed significantly from the control condition, t(376)=2.67, p=0.022, OR=0.44, and t(376)=2.99, p=0.009, OR=0.43, respectively. The difference between the MERs and the control condition was, however, not statistically significant, as were the differences between the three experimental conditions. Accuracy in the control condition was significantly lower than in the three experimental conditions combined, F(1, 376)=9.60, p=0.002, OR=0.33. This confirms our prediction partly: The combination condition scored better than the control condition, as did the refutation condition, and all experimental conditions combined. The MER condition did not score better than the control condition.

We also found a main effect of item on accuracy, χ^2 (2,368)=11.20, p< 0.001, OR= 0.34–0.35, but no interaction effect of condition and item on accuracy, χ^2 (6,3368)= 0.48, p= 0.824. This suggests that not all area items had the same difficulty level, but that the effect of the items on accuracy did not differ between conditions.

In what follows, we discuss each of the three area items in more detail.

Area Item 1—Batteries. Students were provided with two box plots (see Fig. 5) representing the life span of two brands of batteries. The task of the students was to decide which brand they would choose, using the criterion to prefer the brand with the highest chance to get a battery with a life span more than 10 h. The correct response was brand A, as the median was higher than 10 and thus, the chance of having a battery with a lifespan of more than 10 h was probably more than 50 % while for brand B, where the median is 10, the chance was less than 50 %. Giving response B, however, would be a clear indication of the area misinterpretation as in this box plot, the area above 10 was larger than in box plot A. A clear trend was visible in the accuracy rates. The item was solved best by students in the refutation condition (73.2 %), followed by the combination condition (66.1 %), the MERs condition (55.8 %), and the control

Table 1 Accuracy rates (in %) for the three area items, per condition. All items specifically assessed the occurrence of the area misinterpretation. The items are presented and described in the following pages

	Item 1	Item 2	Item 3	Mean for all items
Control (n=45)	48.9	46.7	68.9	54.8
MERs (n=43)	55.8	60.5	86.1	67.4
Refutation (n=41)	73.2	65.9	80.5	73.2
Combination (n=59)	66.1	67.8	86.4	73.5
Mean of all conditions	61.2	60.6	80.9	67.6

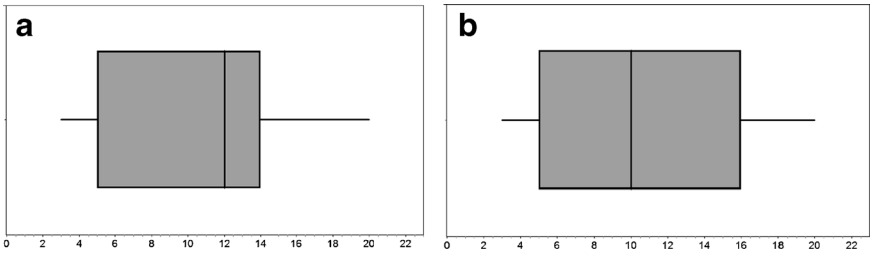


Fig. 5 The box plots used in the "battery problem." Students' had to decide which battery brand they would choose in order to have the highest chance of having a battery with a lifespan of more than 10 h

condition (48.9 %). Although the effect of condition on accuracy was not statistically significant for this item, $\chi^2(3)=6.331$, $p=0.097$, we performed contrast analyses to see whether accuracy was higher in specific conditions. We found that the control condition had significantly lower accuracy than the refutation condition, $\chi^2(1)=5.150$, $p=0.023$, OR=0.35. The difference between the control and the combination condition was almost significant, $\chi^2(1)=3.795$, $p=0.051$, OR=0.49.

Besides looking at the correctness of the responses, we also coded the explanations students provided with their responses. The number of correct explanations was highest in the refutation condition (56.1 % of all explanations in that condition) and lowest in the control condition (37.8 % of all explanations in that condition). The MER (39.5 % of all explanations in that condition) and combination (45.8 % of all explanations in that condition) conditions scored in between. When looking at the explanations of students that showed the area misinterpretation, we see that students who were exposed to the refutational text were less susceptible to this misinterpretation. Students in the refutation condition (9.8 % of all explanations in the refutation condition showed the area misinterpretation) and the combination condition (11.9 % of all explanations in that condition) showed this misinterpretation less than those in the MERs (25.6 % of all explanations in that condition) and control (28.9 % of all explanations in that condition) condition. Some explanations of the students that clearly showed the area misinterpretation were as follows: "Although the median in A is higher, you can see that there is more data below Q2 in box plot A, which means that you have a high chance of getting a battery that lives less than 10 h," "B is more spread out above 10; this gives you a higher chance of a long life span," and "The distance between Q2 and Q3 is higher in B, representing more batteries."

Area Item 2—Glasses. In the second item concerning the area misinterpretation, students were shown the box plot in Fig. 6. Students were asked to determine which group of observations was larger: the group of classes in which fewer than four glass-wearing children were counted, or the group of classes in which more than four glass-wearing children were counted, or whether both groups were equally large. As correct responses, we coded not only "both groups were equally large" but also "you cannot determine this" (because it is possible to construct datasets that result in the same box plot but where the correct response is either one of both sides). Answering that there were more groups with more than four glass-wearing children would suggest that the area misinterpretation played a role, as the area above four was the largest. Looking at the accuracy rates, we found the highest accuracy in the combination condition

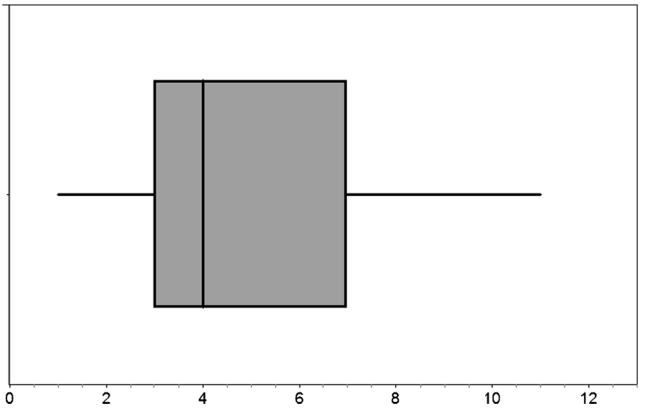


Fig. 6 Box plot used in area item 2. The box plot presents the number of glass-wearing children per group. Students' task was to decide whether there are more groups with more than four glass-wearing children or more groups with less than four glass-wearing children

(67.8 %), followed by the refutation condition (65.9 %), the MERs condition (60.5 %), and the control condition (46.7 %). Although the accuracy rates differed between the conditions, we did not find a statistically significant effect, $\chi^2(3)=5.303$, $p=0.151$. However, contrast analyses showed that the control condition differed significantly from the combination condition, $\chi^2(1)=4.561$, $p=0.033$, OR=0.42. Also, when contrasting the control condition to the three experimental conditions, we found an almost significant effect, χ^2 $(1)=3.625$, $p=0.057$, OR=0.48.

The explanations of the students were, again, more often correct in the three experimental conditions than in the control condition. Specifically, the refutation condition scored best (58.5 % of all explanations in the refutation condition being correct), followed by the combination condition (55.9 %) and the MER condition (53.5 %). In the control condition, this percentage was only 37.8. With respect to the area misinterpretation, students in the experimental conditions scored better than students in the control condition. While in the control condition, 42.2 % of the students displayed the area misinterpretation, this was only 22.0 % in the refutational and combination condition and 25.6 % in the MERs condition. Some typical responses of students include the following: "After four the box plot is the longest," "The distance between Q1 and Q2 is much smaller than the distance between Q2 and Q3," and "You can see that there are more observations after four."

Area Item 3—Matching Histograms and Box Plots. In the third area item, students were provided with the two box plots and two histograms in Fig. 7. Their task was to match the box plots and histograms that depicted the same data distribution. The correct response was that box plot A corresponded to histogram D, while box plot B corresponded to histogram C. When falling prey to the area misinterpretation, however, students would link box plot A to histogram C and box plot B to histogram D. The wide box in box plot A would suggest to these students that a lot of data was situated near the median, while it actually means that the density in this interval was relatively low.

Students in the control condition matched the box plots and histograms correctly in 68.9 % of the cases. Students in the other conditions performed better, with 80.5 % in

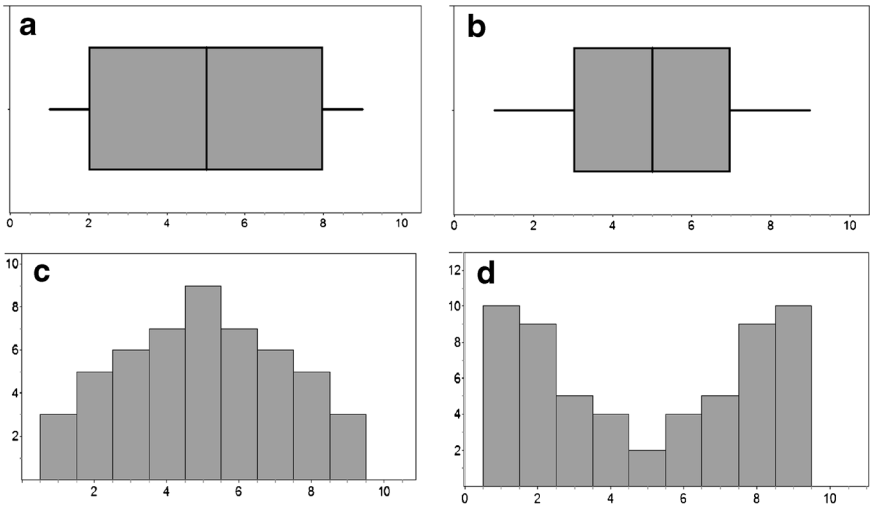


Fig. 7 The box plots and histograms used for item 3. Students were asked to match the box plots and histograms that depict the same data distribution

the refutation condition, 83.7 % in the MERs condition, and 84.8 % in the combination condition. Despite these clear differences, no statistically significant main effect of condition on accuracy was found for this item, χ^2 (3)=5.848, p= 0.119. However, contrast analyses showed again a statistically significant difference between the control and the combination condition, χ^2 (1)=5.124, p= 0.024, OR=0.40. Also, the control condition had statistically significant lower accuracy than the other three conditions combined, χ^2 (1)=4.029, p= 0.045, OR=0.45.

The distribution of explanations for this area item was somewhat different than with the previous ones. Most correct explanations were provided by students in the MERs (72.1 % of all explanations in the MERs condition were correct) and combination (68.5 %) conditions, while the refutation (58.5 %) and control (55.6 %) did less well. The area misinterpretation was displayed most frequently in the refutation (12.6 %) and MER (11.6 %) conditions and least in the control (8.9 %) and combination (6.8 %) conditions. Incorrect explanations that displayed the area misinterpretation are for instance: "In histogram C, most data are in the middle, which you can also see in box plot A," "I looked at the length of the whiskers of box plot B, this is also represented in histogram D by the high bars," and "In histogram D, the low and high values are more frequent, which makes the whiskers in the box plot longer."

Discussion and Conclusion

A first important result is that students in all four conditions scored the same on the box plot knowledge items. This suggests that the four conditions were equally successful in teaching students the theory about box plots. However, students did not score the same in all four conditions on the items that could elicit the area misinterpretation. More

specifically, we found that students in the control condition scored worse than students in the other three conditions, as predicted. Especially, the refutation and the combination conditions scored better than the control condition, with somewhat, but not statistically significant, better results in the combination than in the refutation condition. These results suggest that the use of MERs, refutational text, and their combination worked better to prevent and/or remediate the area misinterpretation than the control condition did. Moreover, it seems that especially, refutational text was effective to improve students' interpretation of box plots and that the use of MERs can enlarge this positive effect. However, in one of the area items, histograms had to be linked to box plots, which was done best by students in the MER condition. This is not surprising, as the link between the histogram and box plot was very explicitly made in this condition. Which instructional technique works best apparently depends on the task that has to be solved.

Our results seem promising, especially when compared to the three studies discussed above. Lem et al. (2013b) did not find any effect of warning students about the fact that box plots can be misleading or of using alternative designs of box plots. Apparently, this small "intervention" was not sufficient for changing participants' interpretation of box plots. The teaching techniques used in the present study, especially refutational text, seemed to be more successful in improving participants' interpretation of box plots. Bakker (2004) did not use pre and posttests, making it difficult to attribute interpretation differences to the used intervention. By using a posttest and four conditions, with this study, we can more confidently match differences in accuracy to specific interventions. Furthermore, our intervention took much less time than the interventions of Bakker, which took several class sessions. Hence, ours is more practical to implement in real classrooms.

We, however, also want to stress that the positive effect of the instructional techniques has to be considered with caution. In neither of the conditions, the area misinterpretation was completely eradicated. The positive effects of refutational text and MERs were at most moderate. Even in the combination condition, in one of the items, only two thirds of the students were able to answer correctly. Despite these drawbacks, we think that the results of this study are very important. First, they show how both MERs and refutational text can successfully be used to remediate the area misinterpretation of box plots. Second, the use of these two teaching techniques separately also showed how each can have a different effect on the solving of different items, showing the multidimensionality of the interpretation of box plots and the differential merits of both techniques. Third, this study shows how the MERs and refutational text can be combined in order to make an even more powerful learning environment. These conclusions seem to be not only an important first step towards the more effective teaching of box plots, but they also show the merits and possibilities of these teaching techniques and hence are a good option to test in the teaching of other graphs and subjects.

Various limitations of this study must be taken into account. A first limitation is that the instructional texts were presented as home study texts. We chose to conduct the study this way because of practical reasons and to raise the ecological validity of letting students study a text as if they were studying outside of a research situation. However, the internal validity may have been compromised somewhat due to this choice: We did not have complete control about several factors, such as the time on task or whether students saw or read the instructional texts of other students who were in other conditions. The fact that home study texts rather than direct teaching were used may also explain why relatively moderate effects were found. A follow-up study could be organized in which students

have to read the text during one of the practical sessions or where the different conditions are taught in the actual statistics class. Although this last option would have various practical difficulties, it is possible that the positive trends and effects that we observed in our box plot test for the MERs, refutation, and combination condition would become larger. Second, most students were already taught about box plots in secondary education. It is not clear whether all these students had the same background knowledge, which could have led to a different effect of the intervention texts on different students. Possible solutions for this potential problem could be to replicate the study in secondary education before box plots are taught or to use a pretest to be able to control for differences in prior knowledge. Third, we found that not all area items in our posttest were equally difficult. However, as we did not find an interaction effect of item and condition, we do not think this has major consequences. Still, it would be interesting to replicate this study with a more elaborated posttest in which more items of different difficulty levels and with different reasoning mechanisms can be compared. Fourth, due to time and curricular restrictions, we were not able to perform a delayed posttest, making it unclear whether the effect of our intervention would persist over time. This could be an interesting question for future research. A final limitation is that we used questionnaires with multiple-choice questions. Besides the fact that guessing cannot be ruled out this way, this method did not allow us to gather much process data. Future research could use verbal protocol methods to get a better insight in the way students reason when solving the box plot test.

There are various implications of our results for education. First, we have once again shown that the interpretation of box plots is not without difficulties. This means that teachers should pay special attention to them and not treat them like self-evident representations. Second, both MERs and refutational text have proven to be able to yield better interpretation of box plots by students. These teaching techniques can relatively easy be incorporated in statistics classes. Third, the combination of MERs and refutational text seemed to yield even better results than the individual use of both teaching techniques. Teachers can use this to optimize their learning materials.

This study has several theoretical implications. First, with respect to MERs, we can say that this teaching technique seems to have a more positive effect on students' interpretation of box plots than a traditional text about box plots, especially when box plots have to be linked to histograms. Second, we can state the same, and depending on the specific interpretation task, even more strongly so, with respect to refutational text. Third, and most importantly, the combination of these two instructional methods seems to be even more positive for students' interpretation of box plots. This creates opportunities for further research, also in other domains than the interpretation of box plots.

Acknowledgments Stephanie Lem holds a post-doctoral fellowship of the Research Foundation - Flanders (FWO). This research was partially supported by grant GOA 2006/01 "Developing adaptive expertise in mathematics education" from the Research Fund KU Leuven, Belgium.

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-The International Journal on Mathematics Education, 41*, 627–636.

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*, 131–152.

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 191–208). Dordrecht, The Netherlands: Springer.

Babai, R., Brecher, T., Stavy, R. & Tirosh, D. (2006). Intuitive interference in probabilistic reasoning. *International Journal of Science and Mathematics Education, 4*, 627–639.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht: Freudenthal Institute.

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 Northland: International Statistical Institute.

Cobb, P., Gravemeijer, K. P. E., Bowers, J. & Doorman, M. (1997). *Statistical minitools [Computer software]*. Nashville & Utrecht: Vanderbult University, TN & Freudenthal Institute, Utrecht University. Retrieved from www.wisweb.nl.

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

Corradi, D., Elen, J. & Clarebout, G. (2012). Understanding and enhancing the use of multiple external representations in chemistry education. *Journal of Science Education and Technology, 21*, 780–795.

Craeghs, B. (2012). *Box plot, quartile plot en offset quartile plot: Verschillende externe representaties, dezelfde interpretatiemoeilijkheden?* [Box plot, quartile plot and offset quartile plot: Different representations, the same interpretation difficulties?] (Unpublished master thesis). KU Leuven, Belgium.

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.

Evans, J., St., B. T. & Over, D. E. (1996). *Rationality and Reasoning*. Hove, England: Psychology Press.

Faul, F. & Erdfelder, E. (1992). *G*Power (version 3.1.7) [Computer software]*. Bonn: Bonn University.

Gill, M. G., Ashton, P. T. & Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology, 29*, 164–185.

Hynd, C. R. (2001). Refutational texts and the change process. *International Journal of Educational Research, 35*, 699–714.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences, 23*, 681–683. doi:10.1037/0003-066X.58.9.697.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometric, 33*, 159–174. doi:10.2307/2529310.

Lem, S., Onghena, P., Verschaffel, L. & Van Dooren, W. (2012). On the misinterpretation of histograms and box plots. *Educational Psycholog, 33*, 155–174.

Lem, S., Onghena, P., Verschaffel, L. & Van Dooren, W. (2013a). External representations for data distributions: In search of cognitive fit. *Statistics Education Research Journal, 12*(1), 4–19.

Lem, S., Onghena, P., Verschaffel, L. & Van Dooren, W. (2013b). The heuristic interpretation of box plots. *Learning and Instruction, 26*(4), 22–35.

Lem, S., Onghena, P., Verschaffel, L. & Van Dooren, W. (2013c). *Experts' misinterpretation of box plots - a dual processing approach.* Manuscript submitted for publication.

Lem, S., Onghena, P., Verschaffel, L., Van Dooren, W. (2014). Interpreting histograms. As easy as it seems? *European Journal of Psychology of Education*. Advance online publication. doi:10.1007/s10212-014-0213-x

McCrudden, M. T. & Kendeou, P. (2012). Exploring the link between cognitive processes and learning from refutational text. *Journal of Research in Reading, 37(S1)*, 1–25.

Palmer, D. H. (2003). Investigating the relationship between refutational text and conceptual change. *Science Education, 87*, 663–684.

Pierce, R. & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal, 25*, 189–205.

Pierce, R., Chick, H. & Wander, R. (2014). Improving teachers' professional statistical literacy. In H. MacGillivray, M. Martin & B. Phillips (Eds.), *Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008–2012* (pp. 295–309). New York, NY: Springer.

Sinatra, G. M. & Broughton, S. H. (2011, August). Understanding the refutation text effect in conceptual change research: Multiple perspectives. In W. Van Dooren (Chair), *The Role of Refutational Texts in Achieving Conceptual Change.* Symposium presented at the biannual meeting of the European Association for Research on Learning and Instruction, Exeter, England.

Stavy, R. & Tirosh, D. (2000). *How students (mis-)understand science and mathematics*. New York, NY: Teachers College Press.

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education, 8*, 951–970.

Tsui, C.-Y. & Treagust, D. F. (2003). Genetics reasoning with multiple external representations. *Research in Science Education, 33*, 111–135.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, England: Graphic Press.

Van den Broek, P. & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22*, 335–351.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D. & Verschaffel, L. (2004). Remedying secondary school students' illusion of linearity: A teaching experiment aiming at conceptual change. *Learning and Instruction, 14*, 485–501.

Waldrip, B. & Prain, V. (2012). Developing an understanding of ions in junior secondary school chemistry. *International Journal of Science and Mathematics Education, 10*, 1191–1213.