The effect of an online course on social sciences university students' understanding of statistics

Vangelis Karagiannakis

Utrecht University

Author Note

The student number of the author is 3705986. Paul Drijvers and Peter Boon had been the supervisors of this research project, which counts for 30 ECTS. In case of publication the target journal is: The International Journal for Technology in Mathematics Education [http://www.tech.plym.ac.uk/research/mathematics_education/field%20of%20work/ijtme/index.htm].

Abstract

Over the past decade there has been an increasing effort of statistics education reform that focuses on statistical reasoning. One means of improving students' statistical reasoning is through emphasizing on the notion of variability. Also, digital tools have shown their potentials to support specific instructional efforts. The subject of the current study is to describe the effects of an online course on nine social sciences student's development of statistical reasoning. The results of the study suggest that students augmented their conceptual development related to the concepts of central tendency and variability by showing a moderate improved level of statistical reasoning. In addition, the online course found to be supportive for students' increasing level of enjoyment when studying and practicing statistics through digital material.

The effect of an online course on social sciences university students' understanding of statistics

Introduction

Statistics education has been the focus of researchers because statistical reasoning is essential in various fields and provides many interesting issues and challenges. Over the last 10 years, there has been a push from statistics education researchers to change the traditional lecture-only format to a more activity-based format. Leaving traditional teaching methods behind in order to promote active learning techniques such as the integration of digital tools can be quite challenging.

One of the main arguments in favor of this reform presented in the literature is that traditional approaches of teaching statistics focus on skills, procedures and computations, which do not lead students to reason or think statistically. Regarding the fact that statistics constitute a main part of social sciences studies, the need for developing statistical reasoning skills becomes a focal point of teaching and learning. In addition, since students who study social sciences probably do not have a solid mathematical background, the selection of digital teaching tools should facilitate their conceptual development and overcome any obstacles in the learning process.

Moore (1990) has positioned statistical ideas like variation or variability at the heart of statistical enquiry as one of the "big ideas" that should underpin statistical teaching. Also, Wild and Pfannkuch (1999), by providing a comprehensive description of the processes involved in statistical thinking -from problem formulation to conclusions-, described the centrality of variability as a fundamental component of statistical thinking. However, variability did not receive much attention in statistics education research. A reason for this is the focus on the measurements of central tendency (Bakker, 2004). The overall goal of this study is to design an online statistics course for university students in social sciences and to evaluate its effects on students' development of statistical reasoning associated with the concepts of central tendency and variability. Furthermore, we intend to describe the extent to which students appreciate the integration of these applets as supplement material to the lectures.

Theoretical Framework

The theoretical framework used in this study focuses primarily on the role of variability in statistics education, on the notion of statistical reasoning, and on the use of digital tools in statistics education.

Variability and its Importance in Statistics Education

Variability is what makes statistics so challenging and interesting and allows us to interpret, model and make predictions from data (Gould, 2004). David Moore, the statistician and former president of IASE and ASA has documented the importance of the notion of variability and its crucial role in both descriptive and inferential statistics. Moore (1990) mentioned that the student must be able to recognize the omnipresence of variability and learn how this variation is quantified and explained. Moreover, Wild and Pfannkuch (1999) have reported the importance of variability as one of the fundamental types of statistical thinking. Many other researchers (e.g., Meletiou-Mavrotheris & Lee, 2002; Reading & Reid, 2005; Reading & Shaugnessy, 2004) have argued that the understanding of the concept of variability can contribute to the development of students' statistical literacy, reasoning and thinking; notions that we refer to in the next section. However, very little research has been conducted on students' understanding of variability (Reading & Shaughnessy, 2004;), despite the central role this concept plays in statistics (Hoerl & Snee, 2001; Moore, 1990).

Variability is interconnected to many core statistical ideas included both in descriptive and inferential statistics and this fact increases the need for students to develop a deep understanding of it. Statistics education researchers (Meletiou & Lee, 2002) have found that students struggle with the concept of variability. Ben-Zvi and Garfield (2004a) have argued that variability is a complex topic to understand and teach, and that its understanding is essential for statistical thinking and reasoning.

Statistical Reasoning

It is argued that statistical literacy, reasoning and thinking constitute the new goals of the statistics education reform for students enrolled in statistics classes, and that these goals are not currently being achieved (Ben-Zvi & Garfield, 2004b). However, there is no consensus among researchers regarding the definitions of statistical literacy, reasoning, and thinking. Garfield, delMas and Chance (2003) made a list that summarizes the current thoughts about the three aforementioned terms. According to this list, statistical literacy is described as the ability to understand statistical information or research results, like organizing data, construct and display tables and work with different representations of data. Within the frame of the notion of statistical literacy, an understanding of probability as a measure of uncertainty is included (Garfield, delMas & Chance, 2003).

Furthermore, Garfield, delMass and Chance (2003) defined statistical reasoning as the way people reason with statistical ideas and make sense of statistical information. This process involves the interpretations based on sets of data, representations of data, or statistical summaries of data. The skills described by statistical reasoning include the ability to connect concepts (e.g., center and spread), or it may combine ideas about data and chance. Ultimately, statistical reasoning means understanding and being able to fully interpret statistical results (Garfield, delMas & Chance, 2003).

Moreover, statistical thinking has been described as the ability to understand why and how statistical investigations are conducted and the "big ideas" that underlie statistical investigations. According to this description, these ideas include the omnipresence nature of variation and when and how to use appropriate methods of data analysis such as numerical summaries and visual displays of data. Further, statistical thinking involves an understanding of more advanced statistical ideas like sampling, how we make inferences from samples to populations, and why designed experiments are needed in order to establish causation. Statistical thinkers are considered able to utilize the context of the problem in forming investigations (from question posing to data collection to choosing analyses to testing assumptions, etc.) as well as to critique and evaluate results of a problem solved or a statistical study.

Ultimately, statistical reasoning is required for decision-making and in particular, variability is what makes decisions in the face of uncertainty so difficult (Garfield & Ben-Zvi, 2005). It is important to mention here that in the 1996 National Assessment of Educational Progress (NAEP) in the United States data that students have weak conceptions of measures of central tendency, and even weaker conceptions of the role and importance of spread in statistical thinking (Zawojewski & Shaughnessy, 2000). For statistical concepts such as measures of central tendency and variability of data, which tend to be difficult or abstract, there are a variety of technological tools to support the development of students' understanding and reasoning (e.g., computers, graphing calculators, Internet, statistical software, and Web applets).

Digital Tools in Statistics Education

It has been argued that supplementing traditional teaching material with tools based on a visual approach and a more active form of learning could improve the effectiveness of teaching (Anderson-Cook & Dorai-Rai, 2001; Cobb, 1992; Marasinghe, Meeker, Cook, & Shin, 1996; Moore, 1997). In order to overcome students' weakness in understanding the processes of statistics, many researchers in statistics education use java applets that allow the simulation of the process of selecting samples and calculating statistics (Garfield & Ben-Zvi, 2008). Such computer tools allow users to dynamically interact with large data sets and to explore different representations in a way that is impossible by hand (Bakker, 2004). In general, it has been shown that the use of applets facilitates students to develop an understanding of statistical concepts (delMas, Garfield & Chance, 1999; Lane & Tang, 2000).

An important advantage of using applets is the interactivity that they offer. The terms interactivity and interaction are widely used in education, but there appears to be no consensus of what they really mean. It is assumed that interactivity is an important – perhaps even essential – attribute of any successful educational technology. Learning is generally more effective (so the argument goes) when the student can control the information exchange. People recall twenty percent of what they see, forty percent of what they see and hear, and seventy percent of what they see, hear and do (Geisman, 1998). In a series of investigations Adams (1992) showed that students using interactive multimedia had a fifty-five percent learning gain over students receiving traditional classroom teaching. Other important findings included that the students learnt material sixty percent faster, and their long-term (30 day) retention was from twenty-five percent to fifty percent higher.

Nowadays, an increased number of statistics educators and researchers use applets with interactive simulations in tertiary education. Both statisticians and researchers recommend that the use of simulations within an introductory statistics course can support specific instructional efforts (Chance & Rossman, 2006). Interactive applets can help students visualize and build a deep understanding of difficult and abstract statistical concepts. Moreover, they offer students the opportunity to see dynamic processes, rather than static figures and illustrations. In addition, the use of interactive applets promote a deeper understanding of statistical concepts by allowing students and designers to pose 'what if' questions and let students test those questions (Garfield & Ben-Zvi, in preparation). These interactive tools are in line with the constructivist view of learning and support students to develop a conceptual understanding of the subject to be learned, rather than knowing isolated disconnected partial bits (Schuyten & Thas, 2007).

Research shows that a good way to use interactive applets is to promote guided discovery learning. In comparison to discovery learning, guided discovery learning has shown much more positive results (Lane & Peres, 2006). More specifically, learning from applets whose design is based on guided discovery learning theory appears to be more effective (de Jong & van Joolingen, 1998). One method of structuring an applet based on the principle of guided discovery learning is by letting students to ascertain answers to specific questions posed beforehand (de Jong, Hartel, Swaak, & van Joolingen, 1996). In that way, students are led to confront discrepancies between their expectations and what occurs when using the simulation. Confronting such discrepancies has been shown to be a potent avenue to conceptual change (Posner, Strike, Hewson, Gertzog, 1982). To this end, online learning environments can offer the potential of providing effective feedback.

Feedback can be used to confront misconceptions and support mindful processing of information and therefore, make learning easier (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Moreno, 2004; Moreno & Mayer, 2005; Timmerman & Kruepke, 2006). To describe the learning effects of feedback, a distinction should be made between feedback that only offers verification of responses and feedback that contains elaboration information providing clues to guide students towards a correct solution (Mason & Bruning, 2000). Regarding the elaboration feedback, there are three different categories; informational, topic-specific and response-specific feedback (Mason & Bruning, 2000). Response-specific feedback, which was used in the design of online materials to improve university students' symbol sense (Tacoma, Drijvers, & Boon, 2011), was proved to be beneficial for informing students about why a given answer is correct or incorrect and what could be done to reach a correct solution. Mason and Bruning (ibid.) suggest that response-specific feedback is the most specific and the most effective of the three. In addition, other researchers claim that immediate response-specific feedback has been proven helpful in making the course content more meaningful to students not only by increasing their motivation but also by allowing to confront their misconceptions quickly (Aberson, Berger, Healy, & Romero, 2003).

Overall, research findings suggest that students learn best when they are exposed to an instructional approach that incorporates both face-to-face and computer-based instruction (Sosa et al., 2011). There is strong evidence that students using computer-based tools in their statistics course demonstrate greater achievement than peers who receive only lecture-based instruction. Although there are many technological tools available, there is still a lack of research on how to best use these tools and how they affect students' learning. The main objective of this study is to design an online course in statistics that aims at students' development of statistical reasoning. Regarding this development, an evaluation of the effects that online activities have on students' conceptual development related to the concepts of central tendency and variability will be performed. As the nature of this study is design-oriented, the focus will also be on the way students improve their understanding about statistical concepts through their work on specific designed tasks.

Ultimately, the current study examines the degree to which students appreciate the use of applets as a tool to learn and practice statistics. The aim of this examination is to gain an insight into students' thoughts and impressions of the online material. Since many research outcomes (Hilton & Christensen, 2002; Alajaaski, 2006; DeVaney, 2010) have contrasting results about the impact that online digital tools have on students' learning experience, the intention is to capture how students react on such a digital teaching intervention. Therefore, the goal of this study leads to the following research questions.

1. Do the integrated online activities enhance social sciences students' statistical reasoning concerning the concepts of central tendency and variability?
2. To what extend do students appreciate the integration of these online materials -as supplement to the lecture-based instructional format- in order to learn and practice statistics?

Methods

Context

The present study is a small-scale explorative research study, in which an online statistics course is designed and field-tested. The digital environment in which the online course was developed is the Digital Math Environment (DME). The practical relevance of this study was to design an online course to supplement the lectures of a freshmen summer course in statistics offered by AUC (Amsterdam University College) to first year students in Social Sciences. The department of Social Sciences organize such a summer course in order to prepare the students, who are coming from the secondary education, in statistics.

The online course was field-tested in two different settings; in the setting of a Statistics course for second year students in Pedagogy and Educational Sciences (Utrecht University), and in the setting of the Basic Research Methods and Statistics I course offered to first year students in Social Sciences (Amsterdam University College). The participants of the study were nine in total. From those nine participants, two were coming from the

Pedagogy department, two from the Educational Sciences department and five from the Social Sciences department. This small number of participants was due to the fact that the students had to volunteer in a busy academic period.

Designing the Modules

The digital environment in which the online course was designed in order to investigate the research questions of the study is the Digital Math Environment (DME). The DME environment has scored highest in the evaluation that has been done by Bokhove and Drijvers (2010). One of the components, which make this digital environment effective, is the potential to provide immediate response-specific feedback. This type of feedback was used in the online course with the aim to inform students whether a given response is correct. In few cases, response-specific feedback was also used to inform students why a given answer is wrong and what could be done to find the correct solution. Such a case is shown in Figure 1.

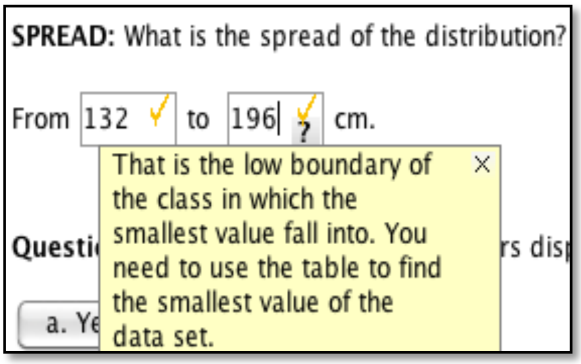


Figure 1. Immediate response-specific feedback

The first part of the study was the design of the two modules included in the online course. Two types of tasks have been designed for developing the two modules; these were experimentation and practice tasks. The main resource for the selection of subjects included in the designed tasks was the textbook *The practice of Statistics in the Life Sciences* (Baldi & Moore, 2012) that was used in the academic course Basic Research Methods and Statistic I.

In the theoretical framework section we referred to the effectiveness of the guided discovery learning technique, a technique that was used to cover the targeting concepts. Initially, the guided discovery learning technique served the research goal of examining students' starting point related to the targeting concepts. Students' thoughts about different graphical representations have been captured through a set of specific multiple-choice questions (hereafter referred to as MCQ) at the beginning of the online course. The initial set of multiple-choice questions (MCQ) is listed below:

1. Does the number of classes chosen in a histogram affect its shape?
2. The histogram of a distribution can be unimodal and bimodal according to the number of classes chosen.
3. Does a stem-and-leaf plot provide enough information about the shape of a distribution?
4. Which measure of central tendency, mean or median, is more resistant measure when outliers are involved in the data set?
5. What is the relationship between mean and median in a right-skewed distribution?
6. In a right-skewed distribution, the first quartile will be farther below the median than the third quartile is above it.
7. Is the standard deviation a resistant measure of spread when outliers are involved in the data set?
8. Does the mean and standard deviation of two data sets provide enough information in order compare the skewness of the related distributions?

Then, based on a hypothetical learning trajectory (see Table 1), in which interactive experimentation and practice tasks were incorporated, students were guided to answer the MCQ for second time. This process offered the possibility to examine the degree to which students augment their conceptual understanding related to the targeting concepts.

Table 1

Hypothetical Learning Trajectory for Experimentation Tasks

	Task #	Mental Activities	Learning Activities
MCQ 1 & MCQ 2	1_4	Visualize how individual values (sample) fall into the bars of a histogram Interact with the applet by changing the number of classes chosen for the histogram and visualize on the screen the changes in the shape of the histogram	Reflect on the observations by choosing the number of classes which best describe the overall pattern of a distribution Describe the shape of the distribution (unimodal/symmetric) Find the class in which the midpoint (center) of the distribution falls into Indicate the spread of the distribution Detect and record outliers as displayed in the histogram
MCQ 3	1_7	Visualize on the same screen (side-by-side) the histogram and the stem-and-leaf plot of the same data set	Observe and compare the shape of the distribution as displayed in the two different graphical representations Find the midpoint and spread of the distribution as displayed in the stem-and-leaf plot Detect and record the outliers as displayed in the stem-and-leaf plot Relate the existence of outliers to their graphical representation as displayed in the histogram

MCQ 4 & MCQ 7	2_4	Interact with the data set of a distribution –through the use of a slider- by changing the value of one individual to be an outlier Visualize on the screen of the applet the changes of the distributions as displayed in a histogram Visualize on the screen the changes in the mean, median and standard deviation while the one individual becomes an outlier	Observe and record the minimum and maximum value that the changing individual can take Reflect on and describe the changes of mean and median while the one individual becomes an outlier
MCQ 5	2_5	Visualize three different distributions on the screen of the applet Observe the values of mean and median for each distribution	Describe the shape of each distribution Compare the values of mean and median regarding the skewness of the distribution Reflect on the comparisons performed and generalize the observations by formulating a rule about the relationship of mean and median with referring to the skewness of the distributions
MCQ 6	2_10	Change the skewness of a distribution –from bell-shaped to right skewed- and visualize on the screen simultaneously the box-plot and the histogram of the distribution	Compare the position of quartiles in relation to the median when the distribution is bell-shaped as well as right-skewed Reflect on and generalize the observations by formulating a rule about the position of quartiles in relation to the median regarding the skewness of the distribution
Variability in Histograms	2_7	Visualize 5 different histograms (multiple display on the same position of the screen)	Judge variability of distributions regarding 5 different histograms Compare bumpiness of distributions as displayed in histograms Compare variability of distributions based on descriptive statistics Compare the concept of bumpiness in histograms with the concept of variability as a measure of spread

Note. The Task numbering has been made according to the number of the module included in the online course and the number of the task within each module.

A specific emphasis was put on the graphical representation of distributions – histograms and box-plots- regarding the measures of central tendency and variability. Further, other graphical tools like dot-plots and stem-and-leaf plots were discussed in order to make connections between the different graphical representations. The tasks that were used in the online course aimed at experimenting with statistical concepts and properties as well as practicing statistical reasoning skills. Table 2 presents the hypothetical learning trajectory on which the design of the practice tasks was based on.

Table 2

Hypothetical Learning Trajectory for Practice Tasks

Task #	Learning Activities	DME features	
1_3	Identify the individuals that a data set describes Indicate how many variables the data set contains Classify which variables are categorical and quantitative	Text answer boxes	Immediate response-specific feedback provided
1_6	Describe the overall pattern of a distribution with referring to the context of the data	Multiple choice questions	
2_3, 2_8	Calculate the median and quartiles for small data sets	Text answer boxes	
2_5	Compare the center and spread of distributions as displayed on dot-plots	Multiple-choice questions	
2_11	Describe and compare the overall patterns of two distributions –graphical representation of box-plot- with referring to the context of the data	Multiple-choice questions	

Note. The Practice Task numbering has been made according to the number of the module included in the online course and the number of the task within each module.

Several possibilities of the DME had been used in order to design the experimentation and practice tasks. Firstly, multiple-choice questions for the MCQ were used in order to guide students form expectations about the concepts included in the tasks. Then, drag-and-drop activities for which students had to match description statements to the appropriate statistical terms were used in order to activate students' knowledge (terminology that was used within the online course). Furthermore, text-answer boxes for having students describe characteristics of distributions such as central tendency and variation were included. In this type of tasks, immediate response feedback was provided to allow students check the correctness of their answers. Moreover, multiple-choice questions in conjunction with text-answer boxes were integrated to allow students answer the MCQ for second time and reflect on their observations by writing what they have found throughout the activities.

Data Collection

Three types of data were collected to investigate the effects of the online activities on students' conceptual development and statistical reasoning. The instruments used were pre-tests and post-tests (see Appendix A) before and after the implementation of the online course and students' work logs in DME. In addition, questionnaires (see Appendix B) were used to address the second research question about students' appreciation of the online activities as supplement to the lectures.

To ensure the validity of the instruments used, both pre-test and post-test included items designed and implemented by other statistics education researchers. The items used in the pre-test and post-test were identical and derived from the CAOS (Comprehensive Assessment of Outcomes in Statistics) test (delMas, Garfield, Ooms, & Chance, 2007). The main aim of the development of this test was the design of a reliable assessment tool that consists of a set of items that students completing any introductory statistics course would be expected to understand. An additional goal of the development of this test was to identify areas that students do and do not make significant gains in their statistical understanding and reasoning after instructional interventions.

The second type of data consists of students' work in DME. The DME saved all the answers given by students in each of the designed tasks. Students' work logs were used in different ways. Firstly, students' prior knowledge related to the topics covered by the online course was examined through the MCQ answers at the beginning of the online course. Secondly, students' work logs on the experimentation and practice tasks were used to investigate the extent to which the designed tasks facilitated students to augment their conceptual understanding.

To answer the second research question about students' appreciation of the online activities as supplement to the lectures, we provided students with questionnaires (see Appendix B). Students evaluated the quality of the online course -instructions to experiment with the applets and the clarity of explanations-, through statements on a 5-point scale. Furthermore, students were supported to express their thoughts in open-ended questions in which they were asked what they like best and least about the online course. To ensure the validity of the questionnaires used, the statements included were selected from the instruments used in a research study, which examined the impact that technological tools (applets) have on students' learning experiences (Street & Goodman, 1998).

Data Analysis

Data analysis focused on students' development of conceptual understanding and statistical reasoning. Firstly, we analyzed the results of pre-tests with the aim to measure students' statistical reasoning before the implementation of the online course. Then, we took a closer look at students' work in DME. Initially, we used students' results in the MCQ in order to explain students' erroneous responses in the pre-test as well as to capture students' initial expectations about the targeting concepts included in the tasks. Further, we analyzed students' responses in the experimentation and practice tasks to examine the conceptual development throughout the activities. Finally, we analyzed the results of post-tests to measure students' statistical reasoning after completing the online course. To compare each individual's answers between the pre-tests and post-tests and the answers between the MCQ and after the experimentation tasks, we used the following patterns:

1. A *correct* pattern represents a correct-correct answer.
2. An *increase* pattern represents a wrong-correct answer.
3. A *decrease* pattern represents a correct-wrong answer.
4. A *wrong* pattern represents a wrong-wrong answer.

As far as the second research question is concerned, we analyzed students' ratings on the statements provided in order to measure the extent to which they appreciate the online course. Also, we analyzed students' responses in the open-ended questions in order to capture students' thoughts about this learning intervention.

Results

The following results relate to the performance of the nine participants of the study.

Pre-tests

In this section, we present the results for each item of the pre-test in order to capture students' statistical reasoning before the implementation of the online course. Aim of our analysis is to detect the concepts in which students had a weak understanding before they proceed to the experimentation and practice tasks. Table 3 presents individuals' answers on each item of the pre-test.

Table 3
Pre-test results

Students	Myrto	Mia	Olive	Nick	Agis	Leo	Irini	Vally	Max	Total Per Item
Item 1	1	0	0	1	1	1	0	0	1	5
Item 2	0	1	0	0	0	1	0	1	1	4
Item 3	1	1	1	1	1	1	1	1	1	9
Item 4	0	0	1	1	1	0	0	1	1	5
Item 5	0	1	1	1	1	0	1	0	0	5
Item 6	0	0	1	1	0	0	0	1	0	3
Item 7	1	1	1	1	1	1	1	1	1	9

Item 8	1	1	1	1	1	0	1	0	1	7
Item 9	1	1	1	1	1	1	1	1	1	9
Item 10	1	0	0	1	1	1	0	0	1	5
Item 11	1	0	0	1	1	1	0	0	1	5
Item 12	1	0	1	1	0	1	0	1	1	6
Total Per Student	7	6	8	11	9	8	5	7	10	

Note. A "0" is used to show a wrong answer, and "1" is used to show a correct answer.

Regarding the results on item 1, we found that four students were not able to describe and interpret the overall distribution of a variable as displayed in a histogram. Considering each individual's wrong answer, we realized that students either could not use the appropriate language to describe the shape, center and spread of the distribution or, they could not recognize and describe the existence of outliers as displayed in a histogram.

Furthermore, results on item 2 show that five students could not recognize and match two different graphical representations of the same data set. The two possible explanations for the detected erroneous responses on item 2 are either a wrong interpretation of the median and quartiles in the context of box-plots or, confusion with how outliers are displayed in box-plots. Students either did not know how to relate the position of quartiles to the bars of the histogram or they were confused with the representation of an outlier as a separated dot that lies outside the minimum value of the data set as displayed in a box-plot. It is confirmed by one of the statistics teachers that students have been told to represent the outliers in a box-plots as the minimum or maximum value of the data set. In contrast, item 3 was proved to be easy for all the students, as we did not find any mistaken response in both the pre-test and post-test. Students could easily match a left-skewed distribution to a variable, which describes the scores of an easy test.

Regarding the results on item 4, 5 and 6, we observed that certain students did indeed have difficulties in interpreting information as displayed in box-plots. Four students could not determine which of two box-plots represents larger standard deviation while the same number of students could not realize that only the quartiles provide estimates for percentages of data. In addition, results on item 6 provided further evidence about student's difficulties in reasoning through the use of box-plots. In particular, six students could not understand the graphical interpretation of the median in the context of the box-plot.

In contrast, working with dot-plots seemed to be easier for most of the students. The results on item 7 show us that all the students could compare groups by considering where the most data are and by focusing on distribution as single entities. In addition, seven out of nine students could compare groups by comparing differences in averages. Also, the same students were able to understand that comparing groups does not require samples of equal size, especially when both groups are quite large. However, there were two students who both did not compare properly the mean of the distributions as displayed in two different dot-plots.

The results on item 10 revealed students' weak conceptions about variability as displayed in histograms. Four students judged lowest variability among five distributions by choosing a statement, which described the lowest standard deviation considering how less difference there is among the high of the bars of the histogram. It is important to mention here that the students, who chose this statement, ignored statements that either described lowest standard deviation based on how close the most values are to the mean or, how small the range is. The misconception that these students hold has been highlighted from their results on item 11. On this item, the four students judged highest standard deviation based on the bumpiness and irregularity of the distributions as displayed in the histograms. Last but not least, the results on item 12 showed that six students could not understand that a distribution with the median larger than mean is most likely skewed to the left.

The analysis of pre-test results revealed that certain students had weak conceptions about the notion of variation. Students faced difficulties to interpret statistical information – central tendency, variability- as displayed in different graphical representations. The main difficulties have been detected in working with histograms and box-plots. These findings are partly relevant to prior research findings, which suggest that histograms are particularly difficult for students to conceptually understand. In particular, researchers (Meletiou & Lee, 2002; Bright & Friel, 1998) have reported that students face difficulties in distinguish the two axes in a display of reduced data such as histogram. This difficulty then makes students to confuse the concept of 'bumpiness' with the concept of variation. Also, students have been found to struggle with the interpretation of the numerical description of a distribution –central tendency, variability- as displayed in box-plots.

DME Work Logs

In this section, an analysis of students' results as saved in DME is presented. The aim of this analysis is not only to examine students' starting point related to the concepts covered by the MCQ but also to explain students' erroneous responses in the pre-test. Students' work logs on questions that had been incorporated in the experimentation and practice tasks have been analyzed in order to show whether and how students improved throughout the activities. Table 4 presents students' answers in the MCQ.

Table 4
MCQ results

MCQ #	Myrto	Mia	Olive	Nick	Agis	Leo	Irini	Vally	Max	Total Per MCQ
1.	1	1	1	1	1	1	1	1	1	9
2.	1	1	1	0	0	0	0	1	1	5
3.	1	1	1	0	1	1	1	1	1	8
4.	1	1	1	1	1	1	1	1	1	9
5.	0	1	1	1	1	0	1	0	0	5
6.	0	0	1	1	0	1	0	1	0	4
7.	1	1	1	1	1	1	0	1	1	8
8.	1	0	1	1	0	1	0	1	1	6
Total Per Student	6	6	8	6	5	6	4	7	6	

Note. A "1" is used to show a correct answer for each multiple-choice question (MCQ), and a "0" is used to show a wrong answer.

Regarding students' results on MCQ1, we see that all the students could understand that the classes chosen in a histogram affect its shape. However, there were three students who could not conceptualize a distribution as displayed in a histogram to be both unimodal and bimodal according to the number of classes chosen (MCQ2). Furthermore, there was only one student who believed that a stem-and-leaf plot did not provide information about the shape of a distribution. In contrast, results on MCQ5 revealed that four students could not relate the relationship between measures of central tendency –mean and median- to the shape of the distribution (skewness). In addition, five students found to be not able to interpret the role of quartiles in the context of a box-plot, with referring to the shape of a distribution. Furthermore, results on MCQ7 show that only one student considered the standard deviation as a resistant measure of variability when outliers are involved in the data set. The rest eight students could understand that the existence of outliers can affect the arithmetic value of standard deviation. Finally, results on MCQ8 present three students who consider the arithmetic values of mean and standard deviation as enough information to reason about the shape of a distribution.

To explain the erroneous responses detected on item 1 of the pre-test, we took a closer look at students' work logs on the experimentation and practice tasks related to the description of variables as displayed in histograms. Regarding the work logs on the experimentation task 1_4 (Figure 2), it is evident that students could describe the symmetry, center and spread of a distribution as displayed in the histogram.

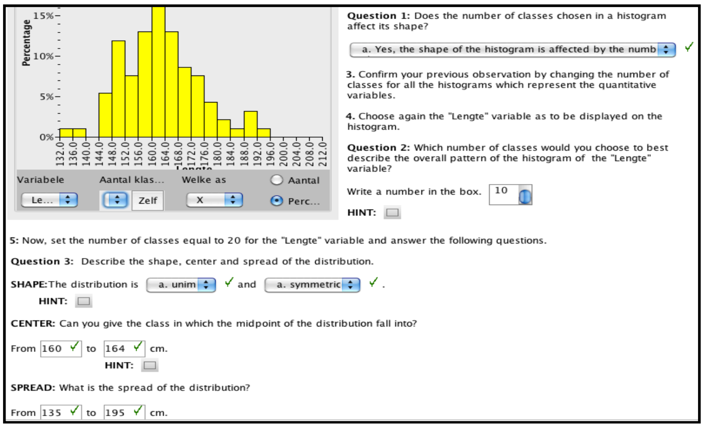


Figure 2. Experimentation task 1_4

Further evidence on students' ability to properly describe the symmetry of a distribution is derived from the work logs on practice-task 2_5 (Figure 3), where no erroneous answers had been given as far as the description of the overall pattern of distributions is concerned. To ensure that students did not simply rely on the provided feedback until to give the correct answer, we analyzed students' trial logs. This analysis revealed that students did not perform any mistakes.

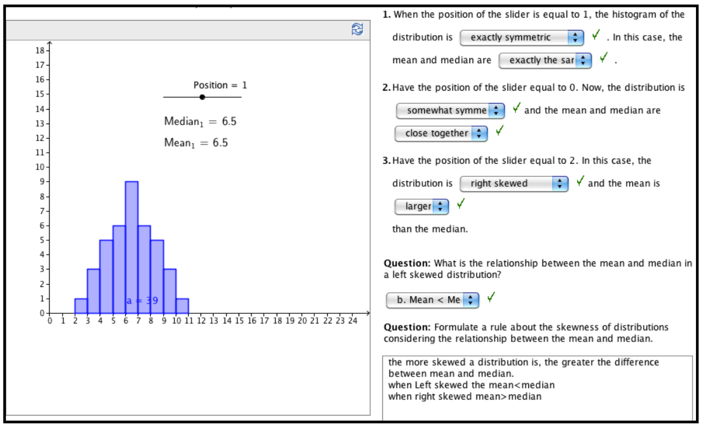


Figure 3. Practice task 2_5

However, students could not easily recognize the existence of outliers as displayed in a histogram. It has been detected in the trial logs of the experimentation task 1_4 (Figure 2) that students had performed more than one mistaken try to determine whether outliers were displayed in a histogram. It is important to mention here that students did not receive feedback in this question, a fact that explains their tendency to choose all the possible answers.

Nevertheless, work logs on the experimentation task 1_7 (Figure 4) show that students progressively have been able to detect outliers by comparing the overall pattern of a distribution through the simultaneous display of different graphical representations. Students had been able to detect and record the outliers as displayed in both a histogram and a stem-and-leaf plot of the same data set.

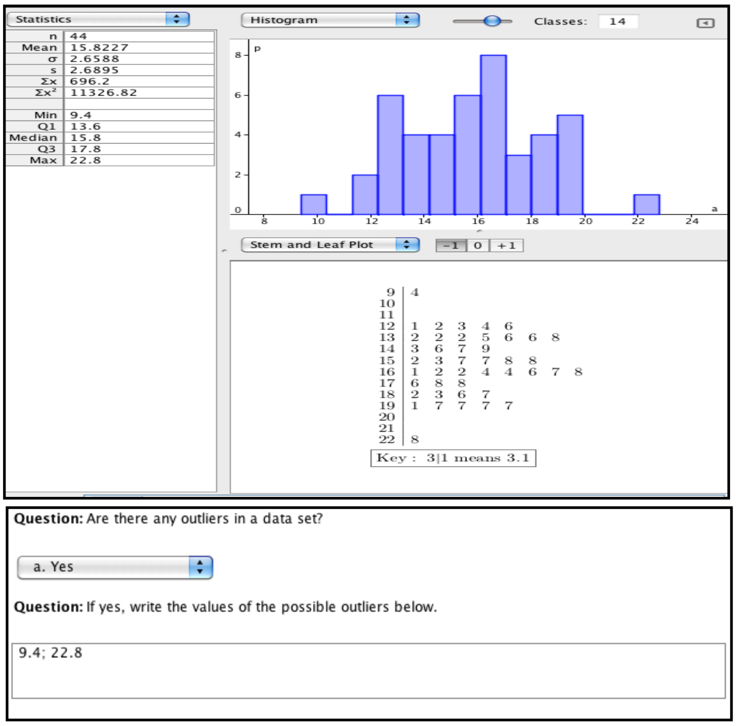


Figure 4. Experimentation task 1_7

Furthermore, we analyzed students' work logs on the experimentation and practice tasks associated with the concept of box-plot as a graphical representation of a data set. This analysis helped us to explain the erroneous responses given on item 2 of the pre-test as well as the erroneous responses detected on the items 4, 5 and 6. Firstly, we have found that five students gave a wrong answer on MCQ6, a fact that supports our observation that students' had difficulties to interpret information as displayed in box-plots. In particular, the wrong answers on MCQ6 show students' limited understanding of the position of quartiles in relation to the median, with referring to the skewness of the distribution. It is important to note here that the same five students had also responded erroneously on item 6 of the pre-test, where they had to interpret the median in the context of the box-plot. These results provided us further evidence about students' difficulties to interpret the position of quartiles in relation to the median, with referring to the context of the data. Nevertheless, students were able to calculate the quartiles of data sets as can it had be seen from the work logs in the practice task 2_8 (Figure 5).

1. Our sample of 15 Aleppo pine needles is arranged in increasing order below:

7.2 7.6 8.5 8.5 8.7 9.0 9.0 9.3 9.4 9.4 10.2 10.9 11.3 12.1 12.8

We have an odd number of observations, so the median is the middle one, the number [9.3 ✓] in the list. The first quartile is the median of the 7 observations to the left of the median. This is the [4 ✓] th of these 7 observations, so Q_1 = [8.5 ✓] cm. If you want, you can use the recipe for the location of the median with $n = 7$:

location of $Q_1 = \frac{n+1}{2} = \frac{7+1}{2} = 4$

The third quartile is the median of the 7 observations to the right of the median, Q_3 = [10.9 ✓] cm.

2. Here are the 18 Torry pine needles, arranged in increasing order:

21.2 21.6 21.7 23.1 23.7 24.2 24.2 25.5 26.6 26.8 28.9 29.0 29.7 29.7 30.2 32.5 33.7 33.7

There is an even number of observations, so the median lies midway between the middle pair, the [9 ✓] th and [10 ✓] th in the list. Its values is M = [26.7 ✓] cm. The first quartile is the median of the first [9 ✓] observations, because these are the observations to the left of the location of the median. Therefore, the first quartile is Q_1 = [23.7 ✓] cm and the third quartile is Q_3 = [29.7 ✓] cm.

Figure 5. Practice Task 2_8

In particular, students could follow the procedure of calculating the median of a data set, as well as the median of the first and second half of the same data set. By looking at students' work logs in the experimentation task associated with the MCQ6 (Figure 6), we found that five students, who gave a wrong answer on MCQ6, showed an increase pattern after the experimentation task. Students through interacting with the histogram of a distribution –change the skewness from symmetric to right skewed- and visualizing on the same screen the changes happening in the associated box-plot have been confronted with their erroneous expectations about the position of quartiles in relation to the median. Also, students have been able to reflect on their observations by formulating a rule about the skewness of a distribution and the position of quartiles.

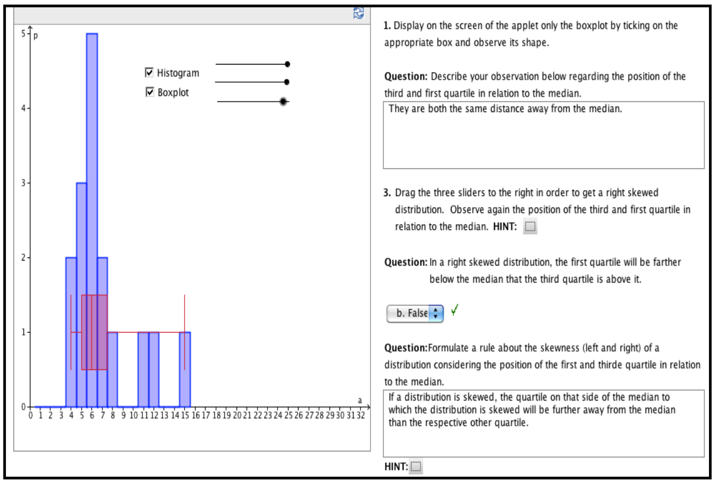


Figure 6. Experimentation task associated with MCQ_6

By following students' work in DME, we have seen that all the students could compare distributions as displayed in dot-plots. This finding is in line with students' results on item 7 and 8 of the pre-test, where only two erroneous responses had been detected.

Regarding students' work logs in the practice task 2_6 (Figure 7), only three mistakes had been recorded when students had to compare the center and spread of the distribution as displayed in dot-plots. These mistakes had been performed from the students, who were not able to compare groups by comparing the differences in averages on item 8 of the pretest.

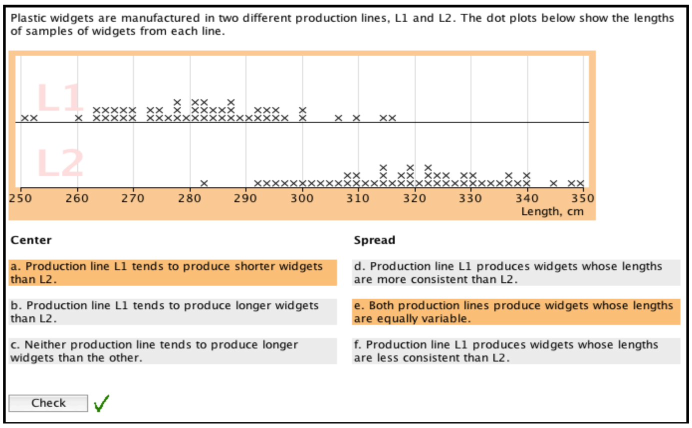


Figure 7. Practice Task 2_6

Regarding students' conceptions about the variability in histograms, we analyzed the work logs of the experimentation task 2_7. The students, who considered bumpiness of histograms as a criterion for the variability of a distribution, again, they performed the same mistake when they had to compare the variability among five histograms. However, students have been confronted with their erroneous expectations throughout the experimentation activity. Students benefited from the comparison of the bumpiness among the five histograms and through comparing the variability using the descriptive statistics have been able to conclude that the concept of bumpiness does not relate to the concept of variability.

Finally, we looked at the answers given on MCQ5 for which students had to compare the mean and median in a right skewed distribution. In this question, we found that the same students, who have responded erroneously on item 12 of the pre-test, considered the mean lower than the median in a right-skewed distribution. However, students through interacting with the shape of the distribution –change the skewness from symmetric to right-skewed- and comparing the mean and median in each case, they have been able to correct their initial wrong answers. Additionally, the same students have been able to generalize their observations by formulating a rule about the relationship between the mean and median with referring to the skewness of the distributions (Figure 6).

Overall, the guided-discovery learning technique was proved to be helpful for students to overcome their erroneous beliefs and develop their conceptual understanding related to the topics covered by the initial set of multiple-choice questions. This is evident by taking a closer look at the patterns that students fell into between the answers given in the initial set of multiple-choice questions and the answers given after the experimentation tasks. Table 5 presents these patterns.

Table 5

Patterns between MCQ and experimentation tasks

Students	Myrto	Mia	Olive	Nick	Agis	Leo	Irini	Vally	Max
1.	0	0	0	0	0	0	0	0	0
2.	0	0	0	1	1	0	1	0	0
3.	0	0	0	1	0	0	0	0	0
4.	0	0	0	0	0	0	0	0	0
5.	1	0	0	0	0	1	0	1	1
6.	1	1	0	0	1	0	1	0	1
7.	0	0	0	0	0	0	1	0	0
8.	0	1	0	0	1	0	1	0	0

Note. A “0” is used to represent a correct pattern and “1” is used to represent an increase pattern.

We see from Table 5, that all the all the wrong answers detected in the MCQ have been corrected after the experimentation tasks. These results support our expectation that students through interacting with data sets and visualizing the changes as displayed in different graphical representations could overcome their initial erroneous beliefs associated with certain concepts.

Post-tests

The patterns that students fell into between the answers given in the pre-test and post-test are presented on Table 6.

Table 6

Patterns between pre-test and post-test

Students	Myrto	Mia	Olive	Nick	Agis	Leo	Irini	Vally	Max
Item 1	0	1	1	0	0	0	1	1	0
Item 2	1	0	1	-1	1	0	1	0	0
Item 3	0	0	0	0	0	0	0	0	0
Item 4	1	-1	0	0	0	1	1	0	0
Item 5	0	0	0	1	0	1	0	0	0
Item 6	1	1	0	0	-1	1	-1	0	1
Item 7	0	0	0	0	0	0	0	0	0
Item 8	0	0	0	0	0	1	0	1	0
Item 9	0	0	0	0	0	0	0	0	0
Item 10	0	1	-1	0	0	0	1	1	0
Item 11	0	1	-1	0	0	0	1	1	0
Item 12	0	1	0	0	1	0	1	0	0

Note. A “0” is used to show a correct pattern, a “1” is used to show a increase pattern, and a “-1” is used to show a decrease pattern.

We see that students' statistical reasoning has increased after the implementation of the online course. Firstly, we observe that all the students had been able to recognize outliers as displayed in a histogram by showing *increase* patterns on item 1. In addition, four out of five students showed an *increase* pattern on item 2, a fact that provide us evidence that students developed their understanding about the relation between two different graphical representations –histogram and box-plot- of the same data set.

Furthermore, the number of the *increase* patterns on item 4 shows us that students improved in working with box-plot and especially, in comparing the variability of distributions. In addition, the two *increase* patterns on item 5 provide further evidence that students understood that the quartiles provide information for percentages of the data set. Students' conceptual development related to the graphical representation of box-plots is also visible by the *increase* patterns on item 6, although two students could not still interpret the median in the context of box-plots.

As far as the item 7 and 8 is concerned, the results of the post-test are not surprising. All the nine students, who were able to compare groups by focusing on distributions as single entities and compare the differences in averages, showed a *correct* pattern. It is not surprising either the two *increase* patterns on item 8. It seems that students through practicing their skills –compare the center and spread of distributions as displayed in dot-plots- and by receiving feedback, were facilitated to realize their erroneous responses in the pre-test by showing an *increase* pattern on the post-test. On item 9, all students show a *correct* pattern, a fact that makes apparent students' ability to understand that comparison between two groups does not require equal size samples.

Moreover, we see that the experimentation task in which students compared the bumpiness in histograms and the variability using the descriptive statistics did indeed had remarkable effects in order students to understand that the concept of bumpiness does not relate to the concept of variability. Regarding the post-test results on item 10 and 11, we can see that there were two and three *increase* patterns respectively. This fact provide us evidence that certain students did indeed got help through their work in the online course to realize that high standard deviation is not related to the bumpiness of the histograms. However, the three *incorrect* patterns confirm prior research findings, which suggest that students difficult overcome their tendency to consider bumpiness as a criterion for high variability. In particular, prior research findings have reported that histograms constitute a stumbling stone to develop students' intuitions about variation, although the strong emphasis that is placed on the instructional design. Finally, we found students who could not relate the relationship between mean and median to the shape of distribution, in the post-test all of them have shown an *increase* pattern by matching a left skewed histogram to a distribution whose mean is lower than the median.

Overall, we observe that students performed better in the post-test in comparison to the pre-test. The experimentation and practice tasks found to have remarkable effects on students' conceptual understanding related to the concepts covered by the instruments used. The interaction with multiple displays –histograms, stem-and-leaf plots, box-plots- which are a transformation of raw data into a entirely new form has been crucial for students' statistical reasoning, for better understanding of variation and statistical ideas related to measures of central tendency. However, findings from the post-test stress the need for helping students improve their understanding of histograms. There were certain students who, when comparing the histograms of two distributions, could not indicate which of the two distributions had more variation. Since the histograms are among the main graphical tools employed in the classroom to assess the shape of distributions, developing students' understanding of spread when visually interpret a distribution displayed in a histogram is necessary to be able to fully grasp the meaning of one of the most important and difficult concepts encountered in statistic classes-the concept of variation.

Questionnaires

Students' appreciation of the online activities had been examined through the questionnaires. Table 7 presents the statements that were included in the questionnaires and Table 8 shows the associated students' results.

Table 7
Statements included in the questionnaires

Item 1	The online environment was well organized and designed.
Item 2	The information contained in the online environment expanded my knowledge of the measures of central tendency.
Item 3	The information contained in the online environment expanded my knowledge of the measures of variability.
Item 4	The explanations of statistical concepts were easy to understand.
Item 5	The online environment included enough instructions on how to work on the activities
Item 6	The online environment included enough instructions on how to work with the interactive applets.
Item 7	The use of interactive applets enhanced my understanding about the measures of central tendency.
Item 8	The use of interactive applets enhanced my understanding about the measures of variability.
Item 9	The use of interactive applets made the process of learning more enjoyable.
Item 10	The use of interactive applets increased my motivation to study statistics through an online environment.

Item 11	The use of interactive applets increased my motivation to practice statistics through an online environment.
Item 12	The feedback that was provided after the answers boxes was helpful to understand my mistakes.

Table 8

Students' Results

Students	Myrto	Mia	Olive	Nick	Agis	Leo	Irini	Vally	Max
Item 1	3	5	4	4	4	3	4	3	3
Item 2	4	3	4	4	4	4	4	3	4
Item 3	4	5	5	5	4	4	4	3	2
Item 4	4	4	5	4	5	3	4	2	5
Item 5	3	5	4	5	4	3	5	4	4
Item 6	3	5	4	5	4	3	5	4	5
Item 7	4	5	5	3	5	4	5	4	5
Item 8	4	4	5	3	4	2	4	4	5
Item 9	4	4	4	4	5	4	5	5	5
Item 10	4	4	5	3	4	4	4	3	2
Item 11	4	5	4	3	4	2	4	4	5
Item 12	5	4	4	4	4	4	5	3	2

We see that the majority of students found the online course as well organized including information that expanded their knowledge related to the measures of central tendency and variability. Furthermore, students found the explanations provided in the course easy to understand and the instructions as good enough in order to work with the designed applets and activities. Moreover, students appreciated the use of the interactive applet as a tool for better understanding of the discussed concepts.

The most positive finding was that the incorporation of the interactive applets provided a more enjoyable process of learning, a fact that might explain the high rates as far as students' motivation to study and practice statistics in an interactive way is concerned. Also, students highly appreciated the feedback provided as a tool to realize their mistakes and rethink their answers. Regarding students' feedback in the open-ended questions, we found comments that highlighted some of the strengths of the online course. Below, we present few of the comments made by the students.

> "Overall it was an enjoyable way to process the basics of statistics. Especially the interactive applets came in handy."

> "You are able to look back over and over again, and you can study in your own speed: there's no teacher who races through the content. Thereby you can view handy extra content and extra examples which help to make you comprehending the content."

"It does not always provide feedback, but when it does it is very helpful. You are immediately encouraged to rethink your answer.
Playing with graphs really helps understanding them."

"The applets where useful as you could immediately see the impact of different data/approaches. This test helped me in the preparation for my BRMS exam."

"The interactive applets make the theory easier to understand. It shows directly how, for instance, outliers affect the standard deviation."

Overall, students appreciated the use of the online material as a tool for studying and practicing statistics next to the regular lecture-only format. Students have reported that they were facilitated to better understand the discussed concepts and to this end, interactivity through the designed applets played a crucial role. Most importantly, students have enjoyed the learning process throughout the online course. This finding is in line with prior research findings (Street & Goodman, 1998), which suggest that online interactive material make the learning process more enjoyable for students.

Conclusion

The first aim of this study was to investigate whether the designed online activities could develop students' conceptual development and statistical reasoning. Explanations for the increased students' statistical reasoning can be sought in the design of the activities included in the online course.

Firstly, the guided discovery learning technique was proved to be efficient for guiding students to study statistical concepts, ideas and properties. The process of answering the same questions twice allowed students to realize their erroneous beliefs, by reflecting on their observations after the experimentation with the designed applets. In the design process, DME features like multiple-choice questions in conjunction with text-answer boxes should be taken into account as they can guide students into thinking about the targeting concepts and record their findings with the opportunity of the teacher's intervention, if necessary. To this end, the DME possibility to save all the work of students played a crucial role. Furthermore, other features like drag-and-drop items can be used for practicing skills from a linguistic point of view. Students could get familiar with the appropriate language and be aware of how to best use it in order to reason about statistical ideas.

The interactivity through the Geogebra applets was found to be beneficial for students. The Geogebra applets succeeded in graphically illustrating statistical concepts by providing students the opportunity to see how these concepts function together as a system. Students were facilitated to visualize statistical principles, which in turn made them easier for students to understand. The integrated applets allowed students to go beyond simply studying and learning about theory to actually seeing its applications, which improves both procedural and declarative form of knowledge.

Another important aspect on which the DME was proved to be very useful is the provision of feedback. By looking at students' work logs, we observed that students did not give up when they gave wrong answers. Instead, we found that the immediate feedback stimulated students to rethink and correct their erroneous answers, a process that reflects the findings of Mory (2003). The positive effects of feedback that we have seen mostly concern verification feedback on the correctness of responses. The observation that the provided feedback played a role in the motivation of students to find the correct solution is similar to the findings of Erev et al. (2006), who regard motivation as one of the most important effects of immediate feedback.

The second aim of this research study was to measure students' appreciation of the online activities as supplement to the lectures. Overall, the online course was rated highly and comments about the applets indicated that they were "helpful" to better understand the discussed concepts. Students were comfortable manipulating the applets as instructed and thought that they added to their learning experience. Another fact that was highly appreciated by the students was the flexibility of having a substantial control of their learning process, by completing the course at their own pace of learning. This outcome is in line with prior research findings from the field of cognitive psychology and constitutes the hallmark of constructivist learning. Most importantly, the online course increased the learner's enjoyment level, a finding that is relevant with prior research outcomes (Street & Goodman, 1998), which suggest that the online course could stimulate students to focus on understanding the information taught. With this increased enjoyment may come a greater propensity to focus on the significant statistical concepts and ideas; and hence to promote learning and understanding.

Discussion

Overall, we have found that the designed online material had a positive impact on students' conceptual development related to the targeting concepts by increasing the level of statistical reasoning. However, the small number of participants does not allow us to generalize this conclusion. In fact, the moderate positive effects that the online course had on students' development of statistical reasoning constitute a baseline on how to best use these digital tools in the future in order to promote a deeper understanding of statistical concepts and ideas. Regarding at the designed applets both as a demonstration tool for the teacher and as an e-learning object on a web-based environment, we are confident about the potential to promote classroom discourse in which students exchange statistical arguments that focus on significant statistical ideas.

Additionally, the designed applets provided a graphic and interactive presentation format, which can enhance semantic elaboration; therefore a future research in whether learning through online material can lead to better long-term retention might also provide more evidence on the usefulness of interactive applets. Further research on the complementary verbal and visual representation in conjunction with interactions through sliders might offer cues on how to reinforce the relationships between the concepts presented by the objects on the screen in order to improve learning and memory. In addition, since only few concepts related to descriptive statistics were included in the online course, an employment of interactive applets that allow the simulations of the process of selecting samples, repeating measurements and experiments could provide further evidence on how to best use interactive applets to teach more advanced statistical ideas associated with inferential statistics.

Moreover, another important aspect of the DME features that might need further research is the provision of effective feedback. Considering the fact that the provided feedback can increase students' motivation, research on how to best incorporate multiple types of feedback should be taken under consideration. As it is stressed successful feedback contains both verification and elaboration (Mason and Bruning, 2002; Mory, 2003) and the employment of more possibilities for elaboration feedback can further enhance the potential of the DME for developing students' statistical reasoning. We believe that through appropriate use of design principles such as the guided-discovery learning technique integrated with interactive applets, useful material can be designed for the development of students' statistical reasoning.

Ultimately, our research finding that the online course increased learners' enjoyment level supports our view that students could focus on significant statistical concepts and ideas and, therefore, to develop students' conceptual understanding. This may also be relevant to an important issue that has not been directly addressed in this study, but which is certainly important in determining the overall effectiveness of the online course interactivity in learning long-term information retention. However, we perhaps need to focus more specifically on understanding a student's individual reaction to the learning process, which in turn could lead to a more rigorous understanding of when- and how- interactive components can be most efficiently employed.

References

Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. J. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology*, *30*, 75–78.

Adams, G. (1992). Why Interactive? *Multimedia and Videodisc Monitor, 10*(3), 20-24.

Alajaaski, J. (2006). How does technology affect students' attitudes towards the discipline and study of mathematics/statistics?. *International Journal of Mathematical Education in Science and Technology, 37*(1), 71-79.

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., … Utts, J. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project College Report*. California, American Statistical Association.

Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., … Witmer, J. (2005). GAISE College Report, *American Statistical Association*. Retrieved June, 20th, 2013, from http://www.amstat.org/education/GAISE/GAISECollege.htm.

Anderson-Cook, C. M., & Dorai-Rai, S. (2001). An active learning in-class demonstration of good experimental design. *Journal of Statistics Education*, *9*. 9-22.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht: CD-β Press.

Baldi, B., & Moore, D. S. (2012), *The practive of statistics in the life sciences*, (2nd ed.). New York, W.H. Freeman and Company.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213–238.

Ben-Zvi, D., & Garfield, J. (2004a). Research on Reasoning about Variability- A Forward. *Statistics Education Research Journal*, (2), 4-6. Retrieved October, 20th, 2012, from http://www.stat.auckland.ac.nz/serj.

Ben-Zvi, D., & Garfield, J. (2004b). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bokhove, C., & Drijvers, P. (2010). Digital tools for algebra education: Criteria and evaluation. *International Journal of Computers for Mathematical Learning*, *15*, 45-62.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12* (pp. 63-88). Mahwah, NJ: Erlbaum.

Chance, B., Ben-Zvi, D., Garfield, J., Medina, D. (2007). The role of technology in improving students' learning of statistics. *Technology Innovations in Statistics Education*, *1*(1), 2007.

Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. Paper Presented at the Seventh International Conference on Teaching Statistics, Salvador da Bahia, Brazil.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, *23*(7), 13-20.

Cobb, G. (1992). Teaching statistics. In L.A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3-43; MAA Notes No. 22). Washington, DC: Mathematical Association of America.

de Jong, T., Hartel, H., Swaak, J., & van Joolingen, W. (1996). Support for simulation-based learning; the effects of assignments in learning about transmission lines. In A. Diaz de Ilarazza Sanchez and I. Fernandez de Castro (Eds.), *Computer Aided Learning and Instruction in Science and Engineering*, (pp. 9-27). Berlin: Springer Verlag.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*, 179-202.

delMas, R., Garfield J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal* *6*(2), 28-58. Retrieved September, 15th, 2012, from http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_delMas.pdf

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning. *Journal of Statistics Education, 7*(3). Retrieved September, 15th, 2012, from http://www.amstat.org/publications/jse/secure/v7n3/delMas.cfm

DeVaney, T. (2010). Anxiety and attitudes of graduate students in on campus vs. online statistics courses. *Journal of Statistics Education, 18*(1), 11-12.

Erev, I., Luria, A., & Erev, A. (2006). *On the Effect of Immediate Feedback*. Retrieved September, 29th, 2012, from http://telem-pub.openu.ac.il/users/ chais/2006/05/.

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22-38.

Garfield, J., Ben-Zvi, D. (in preparation). The Role of Technology in Helping Students Develop Statistical Reasoning. In J. Garfield & D. Ben-Zvi (Eds.), *Developing Students Statistical Reasoning: Connecting Research and Teaching*. Minnesota, Key College Press.

Garfield, J., & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Springer.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92-99. Retrieved September, 15th, 2012, from http://www.stat.auckland.ac.nz/~iase/serj/SERJ4(1)_Garfield_BenZvi.pdf]

Garfield, J., delMas, R., & Chance, B. (2003). Web-based assessment resource tools for improving statistical thinking. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Geisman, J., (1998). Beyond CBT: Interactive video, *Computers and Personnel*, 35-38, 1998.

Gould, R. (2004). Variability: One statisticians view. *Statistics Education Research Journal, 3*(2), 7-16. Retrieved September, 15th, 2012, from www.stat.auckland.ac.nz/serj

Hilton, S. C., & Christensen, H. B (2002). Evaluating the impact of multimedia lectures on student learning and attitudes. *Proceedings of the Sixth International Conference on Teaching Statistics,* ed. B. Phillips, Voorburg, The Netherlands: International Statistical Institute.

Hoerl, R.W., & Snee, R.D. (2001). Statistical thinking: Improving business performance. Pacific Grove, CA: Duxbury.

Lane, D., & Peres, C., (2006). Interactive Simulations in the Teaching of Statistics: Promise and Pitfalls. In A. J. Rossman and B.L. Chance (Eds), *Proceedings of the seventh international conference on the teaching of statistics* (ICOTS 7), Salvador, Bahia, Brazil. Retrieved September, 20th, 2012, from www.stat.auckland.ac.nz/~iase/publications/17/7D1_LANE.pdf

Lane, D. M., & Tang, Z. (2000). Effectiveness of simulation training on transfer of statistical concepts. *Journal of Educational Computing Research, 22*, 383-396.

Mason, B. J., & Bruning, R. (2000). Providing Feedback in computer-based Instruction: What the Research Tells Us. Retrieved September 29th, 2012, from http://dwb4.unl.edu/dwb/Research/MB/MasonBruning.html.

Marasinghe, M. G., Meeker, W. Q., Cook, D., & Shin, T. (1996). Using Graphics and Simulation to Teach Statistical Concepts, *The American Statistician*, *50*(4), 342-351.

Meletiou-Mavrotheris, M., & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal, 1*(2), 22-37. Retrieved September, 15th, 2012, from www.stat.auckland.ac.nz/serj/SERJ1(2).pdf

Meletiou, M., & Lee, C. (2002). Student understanding of histograms: A stumbling stone to the development of intuitions about variation. In B. Philips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society,* Cape Town, South Africa.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*, 99–113.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*, 117–128.

Moore, D.S. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review*, *65*, 123-165

Moore, D. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New* *approaches to numeracy* (pp. 95-137). Washington, D.C.: National Academy Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211-272.

Reading, C., & Reid, J. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40-59. Retrieved September, 15th, 2012, from http://www.stat.auckland.ac.nz/serj

Reading, C., & Reid, J. (2005). Consideration of variation: A model for curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable* (pp. 36-53) Voorburg, The Netherlands: International Statistical Institute.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer Academic.

Reed-Rhoads, T., Murphy, T. J., & Terry, R. (2006). *The Statistics Concept Inventory (SCI)*. Retrieved September, 15th, 2012, from http://coecs.ou.edu/sci/ .

Schuyten, G., & Thas, O. (2007). Statistical Thinking in Computer-Based Learning Environments, *International Statistical Review*, *75*(3), 365–371.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of chance and data. In F. Biddulph & K. Carr (Eds.), *People in mathematics education: Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6-22). Sydney: MERGA.

Sosa W. G., Berger, D. E., Saw, A. T., Mary, J. C. (2011). Effectiveness of Computer- Assisted Instruction in Statistics: A Meta-Analysis. *Review of Educational Research, 81*(1), 97–128.

Street, S., & Goodman, A. (1998). Some Experimental Evidence on the Educational Value of Interactive Java Applets in Web-based tutorials, *Third Australasian Conference on Computer Science Education,* pp. 94-100, Association for Computing Machinery, "Brisbane, Australia"

Tacoma, S., Drijvers, P., & Boon, P. (2011). The use of a Digital Environment to improve First Year Science Students' Symbol Sense. *Journal of Computers in Mathematics and Science Teaching*, *30*(4), 403-428.

Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education*, *55*, 73–104.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, *67*(3) 223-265.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress* (235-268). Reston, VA: NCTM.

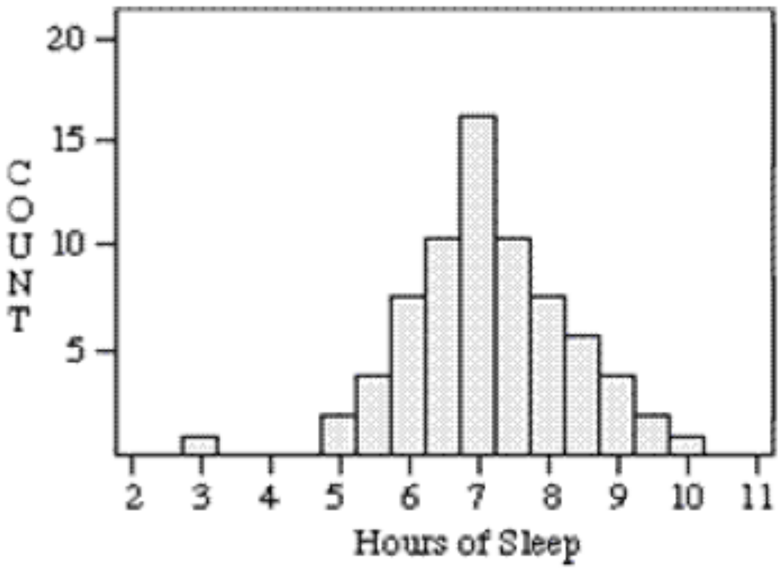
Appendix A

Pre-Test & Post-test

Name:
Email:
Date:

Below there are 15 multiple-choice questions for which only one answer is correct. Highlight or **bold** the answer you think is the correct one for each of the questions provided.

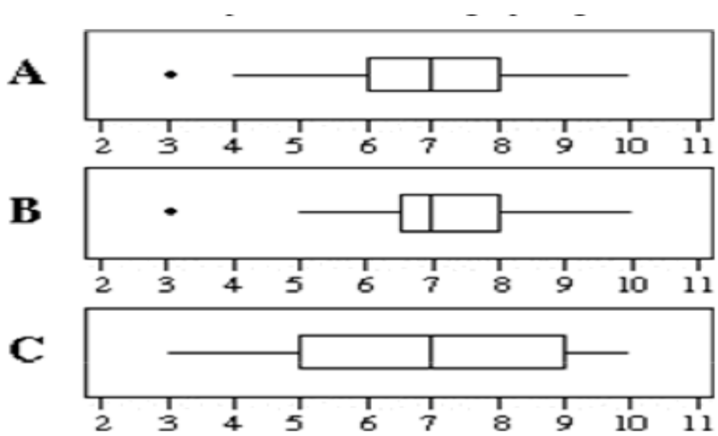
The following graph shows a distribution of hours of slept last night by a group of college students.



1. Select the statement below that gives the most complete description of the graph in a way that demonstrates an understanding of how statistically describe and interpret the distribution of a variable.

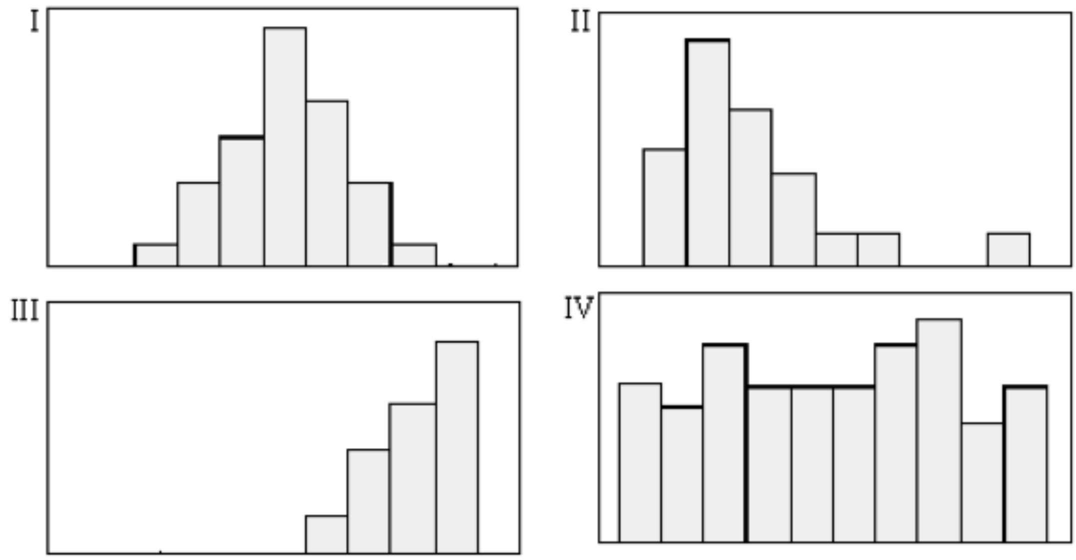
a. The bars go from 3 to 10, increasing in height to 7, and then decreasing to 10. The tallest is at 7. There s a gap between three and five.
b. The distribution is normal, with a mean of about 7 and a standard deviation of about 1.
c. Most students seem to be getting enough sleet at night, but some students slept more and some slept less. However, one student must have stayed up very late and got very few hours of sleep.
d. The distribution of hours of sleep is somewhat symmetric and bell-shaped, with an outlier at 3. The typical amount of sleep is about 7 hours and overall range is 7 hours.

2. Which box plot seems to be graphing the same data as the histogram in question 1?



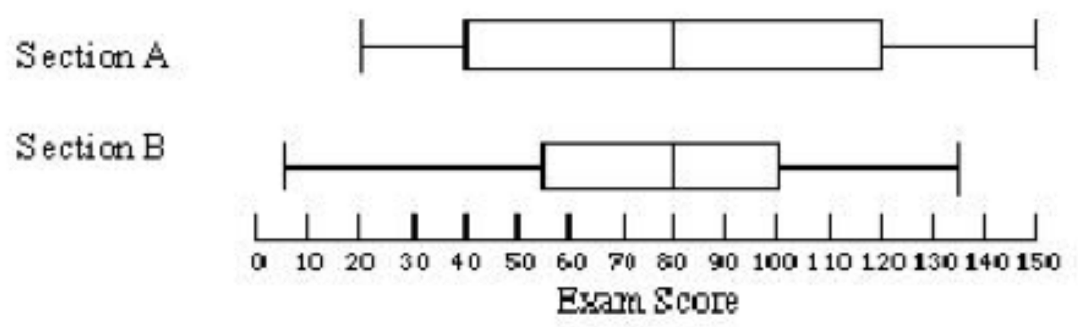
a. Boxplot A
b. Boxplot B
c. Boxplot C

Items 3 to 5 refer to the following situation:
Four histograms are displayed below. For each item, match the description to the appropriate histogram.



3. A distribution for a set of quiz scores where the quiz was very easy is represented by:
a. Histogram I
b. Histogram II
c. Histogram III
d. Histogram IV

Items 4 to 6 refer to the following situation:
The two boxplots below display final exam scores for all students in two different sections of the same course.



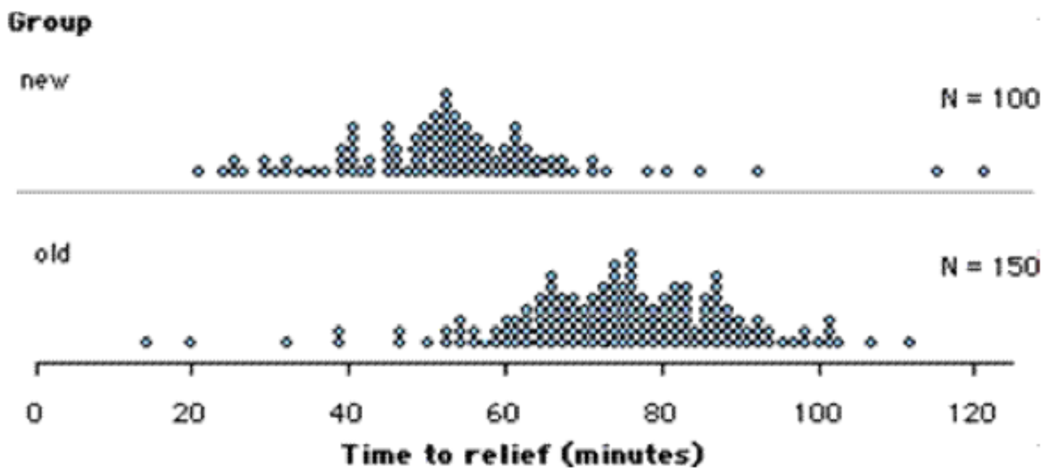
4. Which section would you expect to have a greater standard deviation in exam scores?
a. Section A
b. Section B
c. Both sections are about equal
d. It is impossible to tell

5. Which data set has a greater percentage of students with scores at or below 30?
a. Section A
b. Section B

c. Both sections are about equal
d. It is impossible to tell

6. Which section has a greater percentage of students with scores at or above 80?
a. Section A
b. Section B
c. Both sections are about equal
d. It is impossible to tell

Items 7 to 9 refer to the following situation
A drug company developed a new formula for their headache medication. To test the effectiveness of this new formula, 250 people were randomly selected from a larger population of patients with headaches. 100 of these people were randomly assigned to receive the new formula medication when they a had a headache, and the other 150 people received the old formula medication. The results from both of these clinical trials are shown below. Items 10, 11 and 12 present statements made by three different statistics students. For each statement, indicate whether you think the students' conclusion is valid.

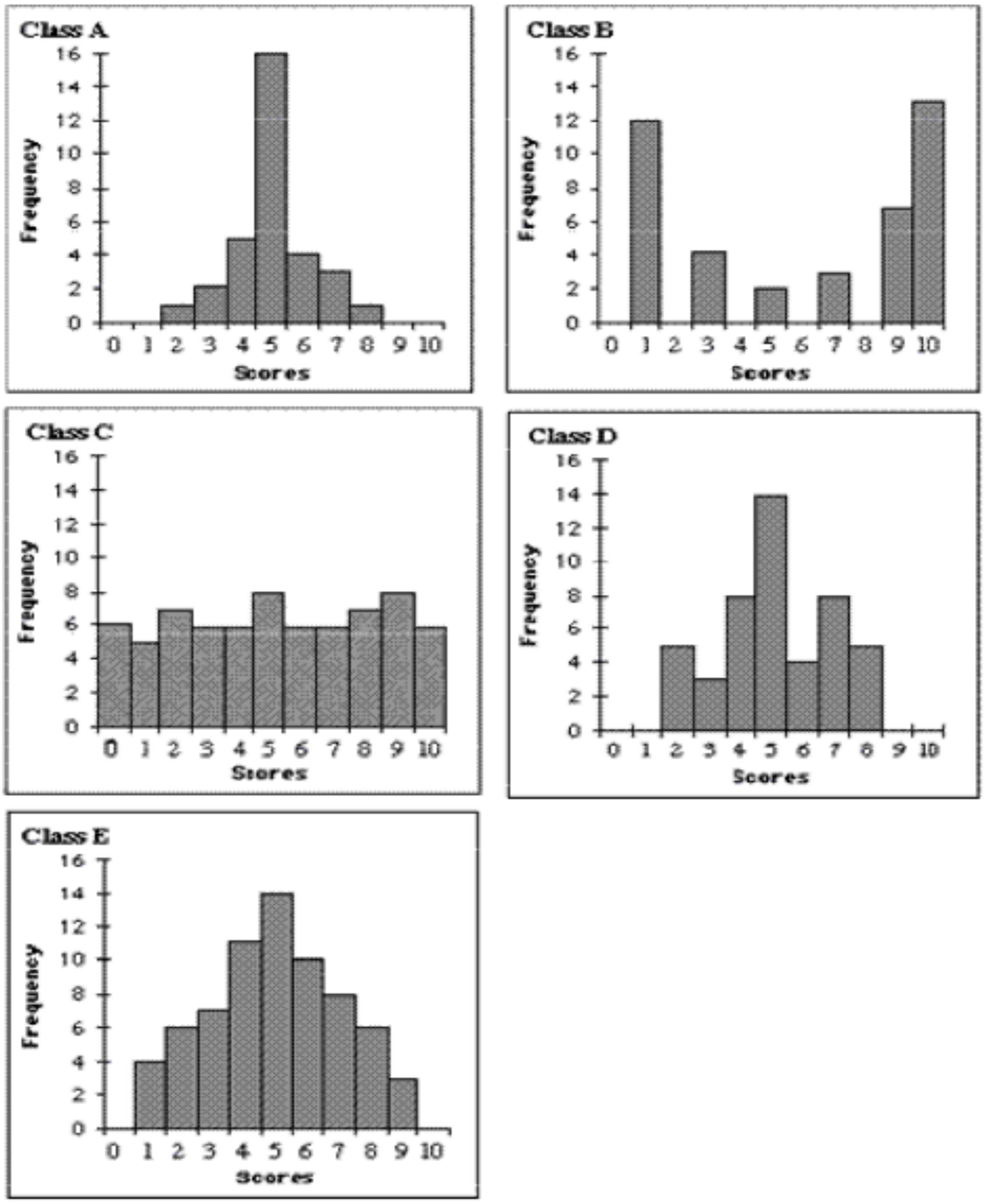


7. The old formula works better. Two people who took the old formula felt relief in less than 20 minutes, compared to none who took the new formula. Also, the worst result –near 120 minutes- was with the new formula.
a. Valid
b. Not valid

8. The average time for the new formula to relieve a headache is lower than the average time for the old formula. I would conclude that people taking the new formula will tend to feel relief about 20 minutes sooner than those taking the old formula.
a. Valid
b. Not valid

9. I would not conclude anything from these data. The number of patients in the two groups is not the same so there is no fair way to compare the two formulas.
a. Valid
b. Not valid

Items 10 to 11 refer to the following situation:
Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.



10. Which of the classes would you expect to have the lowest standard deviation, and why?

a. Class A, because it has the most values close to the mean.
b. Class B, because it has the smallest number of distinct scores.
c. Class C, because there is no chance in scores.
d. Class A and Class D, because they both have the smallest range.
e. Class E, because it looks the most normal.

11. Which of the classes would you expect to have the highest standard deviation, and why?

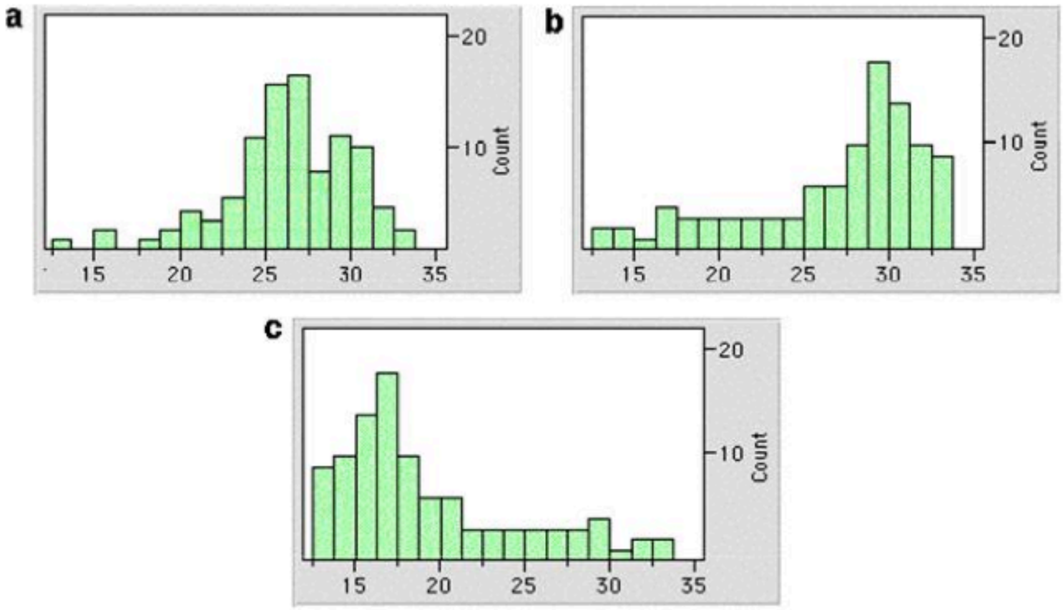
a. Class A, because it has the largest difference between the heights of the bars.
b. Class B, because more of its scores are far from the mean.
c. Class C, because it has the largest number of different scores.
d. Class D, because the distribution is very bumpy and irregular.
e. Class E, because it has large range and looks normal.

A study examined the length of a certain species of fish from one lake. The plan was to take a random sample of 100 fish and examine the results. Numerical summaries on lengths of the fish measured in this study are given.

Mean	26.8 mm
Median	29.4 mm

Standard Deviation	5.0 mm
Minimum	12 mm
Maximum	33.4 mm

12. Which of the following histograms is most likely to be the one for these data?



a. Histogram **a.**
b. Histogram **b.**
c. Histogram **c.**

Appendix B

Questionnaires

The following questions are designed with the intention to gain an insight into your thoughts and impressions about the online environment and the activities that you have worked on. Write you answers in the text boxes and highlight or bold the numbers provided below the statements.

Name:

Current program that you enrolled in. Answer:

Type of secondary education. Answer:

Type of operating system that you logged in the online course.
a. Windows b. Macos

1. The online environment was well organized and designed.

Strongly Disagree				Strongly Agree
1	2	3	4	5

2. The information contained in the online environment expanded my knowledge of the measures of central tendency.

Strongly Disagree				Strongly Agree
1	2	3	4	5

3. The information contained in the online environment expanded my knowledge of the measures of variability.

Strongly Disagree				Strongly Agree
1	2	3	4	5

4. The explanations of statistical concepts were easy to understand.

Strongly Disagree				Strongly Agree
1	2	3	4	5

5. The online environment included enough instructions on how to work on the activities.

Strongly Disagree				Strongly Agree
1	2	3	4	5

6. The online environment included enough instructions on how to work with the interactive applets.

Strongly Disagree				Strongly Agree
1	2	3	4	5

7. The use of interactive applets enhanced my understanding about the measures of central tendency.

Strongly Disagree				Strongly Agree
1	2	3	4	5

8. The use of interactive applets enhanced my understanding about the measures of variability.

Strongly Disagree				Strongly Agree
1	2	3	4	5

9. The use of interactive applets made the process of learning more enjoyable.

Strongly Disagree				Strongly Agree
1	2	3	4	5

10. The use of interactive applets increased my motivation to study statistics through an online environment.

Strongly Disagree				Strongly Agree
1	2	3	4	5

11. The use of interactive applets increased my motivation to practice statistics through an online environment.

Strongly Disagree				Strongly Agree
1	2	3	4	5

12. The feedback that was provided after the answers boxes was helpful to understand my mistakes.

Strongly Disagree				Strongly Agree
1	2	3	4	5

Question 13: What are the strengths and weaknesses of the online course?

Question 14: Think about strengths and weaknesses of the online course and then, think about face-to-face courses.
• What do you miss about face-to-face courses when you are in the online course?
• What do you misses about the online course when you are in a face-to-face course?

Additional Comments: