STUDENTS' EXPRESSIONS OF UNCERTAINTY IN MAKING INFORMAL INFERENCE WHEN ENGAGED IN A STATISTICAL INVESTIGATION USING TINKERPLOTS

ANA HENRIQUES
Instituto de Educação, Universidade de Lisboa
achenriques@ie.ulisboa.pt

HÉLIA OLIVEIRA
Instituto de Educação, Universidade de Lisboa
hmoliveira@ie.ulisboa.pt

ABSTRACT

This paper reports on the results of a study investigating the potential to embed Informal Statistical Inference in statistical investigations, using TinkerPlots, for assisting 8th grade students' informal inferential reasoning to emerge, particularly their articulations of uncertainty. Data collection included students' written work on a statistical investigation as well as audio and screen records. Results show students' ability to draw conclusions based on data, recognizing that these are constrained by uncertainty, and to use them to make inferences. However, few students used probabilistic language for describing their generalizations. These results highlight the need for working on probabilistic ideas within statistics, helping students to evolve from a deterministic perspective of inference to include uncertainty in their statements.

Keywords: Statistics education research; Informal statistical inference; Statistical investigation; TinkerPlots; Probabilistic language

1. INTRODUCTION

Inference and decision making are everyday life activities. Thus, to generalize "beyond data" is relevant to all ages. Informal Statistical Inference (ISI) has become a focus of statistics education research given its potential to help young students in the process of developing important ideas involved in statistical inference, which is a well-known difficult area for older students when formal ideas are introduced (Ben-Zvi, Aridor, Makar, & Bakker, 2012; Makar, Bakker, & Ben-Zvi, 2011; Makar & Rubin, 2009; Zieffler, Garfield, delMas, & Reading, 2008). One of the ideas that stands at the heart of ISI is uncertainty; drawing conclusions about "some wider universe" from observing patterns and trends in data requires students to recognize and articulate inherent uncertainty in predictions beyond the data and probabilistic justification for them (Kazak, 2015; Makar & Rubin, 2009; Rossman, 2008). This implies that developing an understanding of statistical inference requires fostering probabilistic considerations (Manor, Ben-Zvi, & Aridor, 2014).

In line with these ideas, a number of curriculum documents (e.g. Common Core State Standards Initiative [CCSSI], 2010; Franklin et al., 2007; National Council of Teachers of Mathematics [NCTM], 2000) advocate the broadening of both statistics teaching and

probability teaching, and the progressive development of connections between them through the study of inferential statistics, in which one makes inferences that are based on data and qualifies them using probability.

The intention to support learning of the basis of inference and the consistent difficulties students have in reasoning informally about statistical inference, raises questions on how to approach the topic better from a pedagogical perspective (Pfannkuch, 2006). Particularly in Portugal, where middle school students have no experience of informal inference methods, the challenge is to figure out how to approach the topic by choosing tasks that meet the syllabus and, at the same time, support students during the processes of informal inference.

Research in this field begins to offer insights into new approaches to teaching informal inference, particularly involving real data and technology-rich settings, in which students are actively involved in making inferences from samples (Meletiou-Mavrotheris & Paparistodemou, 2015; Prodromou, 2013). Although perspectives and studies have begun to shed some light on important aspects of ISI, further research is needed to build a foundation to guide effective approaches to foster students' inferential practices and their understanding of probabilistic and statistical ideas that underpin those practices. This paper contributes to the field by reporting on the results of a study examining the potential of embedded ISI in statistical investigations. In this study *TinkerPlots^TM* software (Konold & Miller, 2005) was used in order to bring out the informal inferential reasoning of 8th grade students. This objective raises the following research question: What aspects of informal inference emerge when 8th grade students carry out a statistical investigation using the dynamic statistical software *TinkerPlots* for data handling? In particular, how do students articulate uncertainty when engaging in informal inferential practices?

After summarizing some major aspects of the recent discussion on informal statistical inference and the underpinned reasoning, we argue that statistical investigations constitute an ideal context to explore ISI. Then, we describe the methodology of the study, including the task. Finally, after presenting the empirical results of the students' activity in exploring a statistical investigation, we discuss how this learning environment provided an opportunity to support students' development of ISI practices.

2. THEORETICAL FRAMEWORK

2.1. INFORMAL STATISTICAL INFERENCE AND INFORMAL INFERENTIAL REASONING

Informal statistical inference is a term used in literature to refer to students' work on precursor forms of statistical inference, from the starting point of raising questions about data sets through to formal inference (Watson, 2008). It is also described as a reasoned but informal process involving making data-based generalizations and expressing them with uncertainty (Makar & Rubin, 2009). For these authors, three key principles are fundamental to characterize ISI, where the first is particular to the inference process and the other two are specific to statistics: (i) generalizations that extend beyond the description of given data; (ii) the use of data (rather than anecdote) as evidence for generalizations; and (iii) the use of probabilistic language for describing the generalizations, including reference to levels of uncertainty about the conclusions drawn.

The importance given to the ability to draw conclusions beyond the data makes ISI a focal objective of statistical reasoning. Combining several perspectives presented in studies focused on informal statistical inference and the underlying students' reasoning,

Zieffler et al. (2008) define informal inferential reasoning (IIR) as "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (p. 44). This reasoning involves taking into consideration several related ideas, namely: (i) properties of aggregates instead of individual cases; (ii) sample size and its influence on the accuracy of population estimates; (iii) bias control; and (iv) signal, distinguishing between claims that are always true and claims that are often or sometimes true (Rubin, Hammerman, & Konold, 2006).

In connection with this, students are expected to develop an understanding of sampling, that is, to develop awareness of the need for sampling and an ability to evaluate the appropriateness of sampling methods and sample size in sample-based claims about unknown populations. However, students face some difficulties with sample and sampling concepts given their complexity. This requires the coordination of several other notions such as variation, spread, distribution, randomness, and representativeness (Ben-Zvi et al., 2012; Watson & Moritz, 2000). In fact, students are required to understand the importance of random sampling to enable generalization of results from a sample to the wider population, supported by probabilistic justification. Similarly, an understanding of sample variation and the degree of its uncertainty in a real situation is a prerequisite for decision making regarding populations (Prodromou, 2013; Rossman, 2008). Rubin et al. (2006) argue that students are capable of making inferences from data when they see distributions as aggregates. However, students usually show difficulties with such "aggregate thinking", presenting other data perspectives that are considered precursor forms of the aggregate vision. Konold, Higgins, Russell, and Khalil (2015) describe these and demonstrate that students may see data as Pointers, when referring to context instead of data; as Case Values, when referring to values of attributes; or even as Classifiers, when focusing on the frequency of particular data.

In informal statistical inference, a common underlying reasoning process involves "assessing the strength of evidence against a claim" (Rossman, 2008, p. 7) based on observed data. However, studies of students' reasoning about sampling reveal they appear not to use this common reasoning naturally when making a statistical inference because they struggle with the concept of statistical uncertainty, an understanding which is essential in making predictions and judgments about the patterns and trends identified in data (e.g., Ben-Zvi et al., 2012). As documented by these authors, "when faced with the uncertainty of making an informal inference from a sample of data, students may initially articulate strong deterministic statements" (p. 924). Because an overemphasis on deterministic ways of reasoning may inhibit students' ability to develop sound reasoning in probability-laden situations (Nilsson, 2012; Rossman, 2008), the development of probabilistic language in the sense of "statistical tendency and/or level of confidence or uncertainty in a prediction" (Makar & Rubin, 2009, p. 87) is fundamental in reasoning and making decisions based on uncertainty data. Probabilistic language also constitutes an important tool to look at ideas of uncertainty. Recent studies providing analysis of students' expressions of uncertainty in the context of drawing inferences informally show how students deal with such uncertainty. For example, Braham and Ben-Zvi (2015) analyzed 5th grade students' expressions of uncertainty while exploring sampling distributions in a *TinkerPlots* inquiry-based learning environment; they found that such articulations "were shaped by two different views in the way they observed and manipulated the sampling distributions: (1) a move from a global to a probabilistic view, and (2) a move from a local-deterministic to a quasi-probabilistic view" (p. 57).

Assuming that expressing themselves in probabilistic terms is challenging for students, Makar and Rubin (2009) argue that "probabilistic language can be any language

appropriate to the situation and level of students to suggest uncertainty in a speculated hypothesis, that a prediction is only an estimate, or that a conclusion does not apply to all cases" (p. 87). Accordingly, Ben-Zvi (2006) argues in favour of an integration of both informal inference and informal argumentation when aiming to develop students' statistical reasoning in rich learning contexts. Argumentation, employed as a pedagogic approach, challenges students to develop an evidence-based argument in their responses (Fielding-Wells, 2014). In so doing, students will also develop probabilistic language to articulate the degree of certainty embedded in generalizations, which is an important scaffold to developing statistical reasoning (Prodromou, 2013). Students' expressions of uncertainty are also shaped by particular features of the activity structure in which they record and represent the results of their experiments (Ainley & Pratt, 2014). While uncertainty is often expressed qualitatively in informal settings, in later schooling students increasingly quantify uncertainties in informal statistical inferences through confidence levels and margins of error (e.g., Dierdorp, Bakker, van Maanen, & Eijkelhof, 2012).

2.2. DESIGNING TASKS TO SUPPORT ISI

References to ISI appear in curriculum documents (e.g. CCSSI, 2010; Franklin et al., 2007; NCTM, 2000), which recommend that students, as early as in middle school, must be given opportunities to participate fully in statistical processes in order to understand the purpose and utility of data for making sense of the real-world (Makar, et al., 2011; McPhee & Makar, 2014; Prodromou, 2013). In particular, students must engage in posing their own statistical questions (hypotheses) about a meaningful phenomenon, designing and employing a plan to collect appropriate data, selecting adequate graphical or numerical methods to analyze the data, and finally drawing data-based conclusions and inferences connecting the interpretation of results with the original questions (Franklin et al., 2007; McPhee & Makar, 2014). Hence, a possible approach to foster the emergence of students' inferential practices is to embed those processes in a data analysis cycle, like the Pose, Plan, Data, Analysis, Conclusion (PPDAC) cycle (Wild & Pfannkuch, 1999), treating it as an inferential process. Due to their nature, statistical investigations often provide a distinctive context to involve students (even the younger ones) in fundamental components of informal inference, such as decision-making and prediction (Makar & Rubin, 2009; Watson, 2008). For teachers, the use of investigations also provides knowledge that can be used in the design, implementation, and assessment of instruction in statistics and data exploration (Henriques & Oliveira, 2013), because such instructional activity incorporates domain-specific knowledge that targets students' statistical reasoning.

In an attempt to articulate the classroom practices that prevail on a daily basis and the opportunities leading to ISI that those practices represent, Leavy (2010) advocates the use of statistical investigations and provides a conceptual framework for designing tasks supporting the underlying reasoning (Figure 1). According Leavy, such tasks have characteristics in line with those that are considered essential to the activities supporting a focus on inference: to use samples for reasoning about population characteristics and to compare samples of data for reasoning about possible differences between populations. In addition, for Leavy, the selection of tasks that support informal reasoning may also be informed by the degree to which such tasks require students: to use previous knowledge insofar as it is available (Zieffler et al., 2008); to give justifications for generalizations based on evidence (Makar & Rubin, 2009; Zieffler et al., 2008); and to use probabilistic

language on the description of generalizations while referring to uncertainty levels on conclusions (Makar & Rubin, 2009).

Purpose of the investigation	Action on Data	Nature of Statistical activity	Feature of Tasks
Draw conclusions about relationships between the characteristics of groups of observations (*inferential statistics*)	→ Looking beyond the data	Using samples to reason about populations Comparing samples of data to reason about possible differences between populations	Tasks that utilize prior knowledge Tasks that require use of evidence to support generalizations Tasks that draw on use of probabilistic language

Figure 1. Guiding features for design and selection of tasks to support IIR (Leavy, 2010, p. 48)

Currently, students and teachers have access to an increasing range of technology, in the form of educational software such as *TinkerPlots* (Konold & Miller, 2005), in order to explore ISI in rich and meaningful contexts, including through the broader process of statistical investigation (Ben-Zvi et al., 2012). Several studies on the benefits provided by dynamic statistics learning environments recognize *TinkerPlots* as a useful tool to support statistical reasoning and investigations and for making inferential reasoning accessible to young students (Ben-Zvi et al., 2012; Fitzallen & Watson, 2014). Fitzallen and Watson reported that *TinkerPlots* enables students to generate a variety of plots that appear meaningful to them and to use these plots effectively in drawing their conclusions from data. The dynamic nature of this software also facilitates students' reasoning processes involving moving back and forth between formulating hypotheses and constructing plots making sense of the data by manipulating graphs and instantly seeing the results. In this way, when working with *TinkerPlots*, students have the opportunity to develop what English (2014) refers to as meta-representational competence, as they create and re-create a variety of representational forms, and critique and compare the adequacy of such representations to get evidence from data. Findings from Ben-Zvi (2006) also suggest that students use *TinkerPlots* not only as a representation tool, but also as an argumentation tool to express their ideas to others. These and other recent studies (e.g., Braham & Ben-Zvi, 2015) illustrate how the use of such software, in combination with appropriate curricula and instructional settings introducing ISI, leads to new insights in students' reasoning about uncertainty and may support them to develop a strong conceptual base on which later to build a more formal learning of inferential statistics.

3. THE STUDY

3.1. BACKGROUND

This study is part of a developmental research project (DRP), conducted by the two authors, aiming to produce and investigate a sequence of tasks oriented towards the development of basic education students' statistical reasoning, using *TinkerPlots* software. The project involved a group of 11 Portuguese basic education mathematics teachers (grades 5-9). It took place between November and June of the 2013/14 school

year, within a strong collaborative environment among researchers and teachers and among teachers themselves. The teachers were co-responsible for designing and discussing the tasks, conducting experiments in the classroom and reflecting on the whole process. These teachers were highly motivated to learn how to use an instructional approach in line with reform curricular trends in school statistics (Franklin et al., 2007; NCTM, 2000).

3.2. PARTICIPANTS AND METHODS

Adopting a Design Research perspective (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), this DRP followed a teaching experiment design where a sequence of tasks and classroom implementation conditions were improved through iterative cycles (Cobb, Zhao, & Dean, 2009). Several methods were used to collect data in order to enable not just the understanding of important aspects of students' statistical reasoning, but also the collection of artifacts that would enable the review of the instructional process (Stephan, Bowers, Cobb, & Gravemeijer, 2003), particularly the tasks to be proposed in subsequent cycles.

The results reported here come from students' activity on a statistical investigation applied in the study's first cycle, carried out during the second term of the school year in an 8th grade class (ages 13-16 years old) with 30 students (20 boys and 10 girls, who are subsequently referred to by fictitious names). The class teacher was one of the participants in the aforementioned DRP. The data were collected by the class teacher together with a colleague and with the help of a research fellow. They included students' written work (coded as WW), screen recordings (SR) of their work with *TinkerPlots* using *AutoScreenRecorder 3.1 Pro* (Wisdom Software Inc., 2013) and audio records (AR) throughout the lessons. In an initial phase, the co-authors identified examples of the students' work on the task that might provide interesting and specific insights on the characteristics of ISI that emerged in their work. Those examples included students' answers to the questions proposed in the task, namely the claims they made to predict the population characteristics or the existence of differences between two populations, plots they created in *TinkerPlots,* and their explanations to validate those claims. Thereafter, the selected examples, sometimes articulated to complement each other, were qualitatively and interpretatively analysed, using Makar and Rubin's (2009) framework, to show how students: make generalizations beyond the data; use data as evidence for generalizations; and use probabilistic language for describing the generalizations. This analysis focuses particularly on the students' articulation of uncertainty when engaged in informal inferential practices.

3.3. LEARNING CONTEXT AND TASK

The teachers who were involved in the DRP and the authors jointly selected and adapted or produced tasks and materials for a sequence of lessons aligned with the Statistics Reasoning Learning Environment principles defined by Garfield and Ben-Zvi (2008). Two preliminary tasks (Tasks 1 and 2) were applied to build the skills students would need for conducting a statistical investigation and for making inferences, as they were not familiar with this kind of activity. *TinkerPlots* was used in every task as a tool for handling data because it is easy to use and provides a dynamic learning environment to support the development of students' statistical reasoning. Students could create and explore the potential of different graphical representations of their choosing. They could also decide which statistical measures they would use to analyze the data, via the drag-

and-drop facility or a set of informal operators (e.g., separate, stack, and order). Taking advantage of the software's tools, students could also engage in goal-oriented activities that could lead to further insights in data analysis. Tasks 1 and 2 engaged students in exploring real data using *TinkerPlots*, while developing key statistical concepts of the syllabus as well as their reasoning. Because this was the students' first contact with the software, Task 1 was structured and oriented by a set of questions, which enabled students to interpret contexts, to explore data (available in *TinkerPlots* database) in several ways and to use various representations (which were discussed in terms of their appropriateness as an evidence provider). Finally, they used their knowledge of the data to answer one initial question and to make predictions. Task 2 involved students in comparison of distributions using simulation provided by the *TinkerPlots* tools.

A classroom culture that promotes inferential reasoning requires an environment where students are encouraged to address problems using collaborative norms and formal or informal statistical concepts. It also requires a context of complex problems where they can face conflicts with their knowledge and their beliefs about the world (Makar et al., 2011). In all lessons, students worked both in autonomous pairs and participated in whole class discussions, in a classroom culture that supported questioning and the use of evidence for their claims.

This study focuses on Task 3 – *The human body: a study in school* (see appendix), which was the last one of the first cycle of the teaching experiment. This task engaged students in a statistical investigation, aiming to involve them in all phases of the investigation (Wild & Pfannkuch, 1999). This investigation included a set of statistical and probabilistic ideas and knowledge: understanding the need for data (variables, data collection methods) and their influence on conclusions (bias, measurement errors); distribution and variability (building and interpreting different graphical representations and statistical measures, comparing distributions); sampling (collecting samples, obtaining estimates, factors affecting the accuracy of inferences; data noise as a result of the sampling process). As statistical inference is not a learning objective in the basic education mathematics syllabus, it was not a topic for students to study. However, the task also included a set of questions that embody the IIR components described in literature (Makar & Rubin, 2009; Zieffler et al., 2008) and therefore enabled the students to investigate meaningful aspects of the statistical inference practice, in an informal way. For example, the task created conditions for students to use their intuitive or prior knowledge of fundamental concepts and language to make claims to predict, without using formal statistical methodology, the population characteristics (shape, center or spread) or the existence of differences between two populations based on the similarities or differences shown in samples. It also challenged students to give justifications based on evidence, and because they had to explain their reasoning, the task uncovered the arguments and justifications of their predictions and analyses. All this work was strongly supported by the exploration of real data through various representations, facilitated by the statistical learning dynamic environment of *TinkerPlots* (Ben-Zvi, 2006).

4. RESULTS: THE EMERGENCE OF STUDENTS' INFORMAL INFERENCE

As the students progressed through the course of the statistical investigation supported by *TinkerPlots* they constructed multiple and purposeful plots to enhance their understanding of data, and to support them to make inferences about the characteristics of all the students in the school. Examples of these plots and the use students made of them, as well as excerpts of students' claims and explanations, were selected to infer and illustrate which aspects of students' informal inference emerge from that activity and

their articulations of uncertainty when engaged in these practices. To enable students to experience ISI we expected them to undertake three actions, which structure this section: (1) making generalizations that extend beyond the given data; (2) using data as evidence for generalization; and (3) using probabilistic language to describe generalizations. All these actions involve drawing on the underpinning ideas about uncertainty.

4.1. MAKING GENERALIZATIONS BEYOND THE DATA

According to the authors referred to earlier, IIR involves making generalizations that go beyond the data. Such inferential components emerged at different moments of the students' work on the task and with different foci.

The task was introduced with an initial question: "How could you characterize the middle school students in your school regarding some of Vitruvius' measures such as height, foot size, and arm span?" This question raised an interesting discussion in the larger group, where it was noted that the students had a perception that an inquiry of the whole population under study (middle school students from 12 to 14 years old) was not necessary in order to get to know the specific characteristics. Several proposals emerged associated with the idea of sampling and surveying as opposed to census: "Within middle school, with the possibility of selecting a mixed group of students from each class" (WW; Jade & Dino), "We'd better conduct a survey" (WW; Ana & Edu), or "I think that by choosing a class [from each] of the 3 years we could have a good idea of [their] characteristics. Conducting a survey and [because] doing a census will take more time" (WW; Edir & Selma). Although understanding the utility of the inferential process, the students were not unanimous regarding the sampling method to be used to extrapolate the results. However, the majority of them mentioned that randomness was vital to ensure the sample representativeness, as shown in the following answers: "To randomly select [students] from 7th to 9th grades" (WW; Berto, Bina, & Remi) and "To choose randomly students from different classes from each of all grades in middle school" (WW; Ana & Edu). Such an approach for selecting students from different school grades, mentioned by the majority of students, seems to be an indicator that they recognized data noise as an aspect to be taken into account when selecting the sample, in order to avoid any bias in the sample. Another group of students, when giving their opinion on sample size, said that "Our class [of 30 students] is a good example because students have different heights" (WW; Alex & Alfeu). This shows an understanding of the importance of data noise when drawing meaningful conclusions on the population. It also shows an understanding that such a factor overlaps with sample size.

For convenience, students collected data regarding height, foot size, and arm span in their own class. After that, they were invited to ask questions about the phenomenon under study and to predict how their questions might be answered, at first without providing them with data. This approach aimed to later confront students' anecdotal answers with data-based evidence, pushing them to understand the need for data.

Some students did not understand that these questions were intended to help them answer the initial question. Hence, they focused their attention on individual case proprieties, or on an attribute value, instead of seeing the data as an aggregate. At this point, students saw data as "case values" or "classifiers" and asked questions such as "What is the biggest value for the boys' arm span?" (WW; Willy & Fausto), "Does the biggest foot size match the tallest student of the class?" (WW; Berto, Bina, & Remi) and "Which is the mode of the girls' heights?" These are examples of deterministic questions that do not lead students to draw conclusions beyond the data. Other students, however, were able to ask questions such as "What is the average height of students in the class?"

(WW; Manel, Sá, & Paulo) or "Do boys tend to be higher than girls?" (WW; Tadeu & David). These questions show students' attempts to find signals and generally to describe the data focusing on aggregate proprieties and to take into account data noise. These last questions already show some statistical characteristics central to informal inference, including elements of uncertainty. Thus, these students were beginning to adopt a statistical perspective of trend, although still limited to the properties of the class. Regarding their predictions, some students justified them based on guessing and simple observation: "Based on [the observation of] my colleagues, I think that the average height of the students [boys] is 165cm" (WW; Manel, Sá, & Paulo). At this stage, such findings are not surprising, as students collected their own data but did not explore them, and consequently this activity might have diverted their attention from using the data inferentially.

After exploring the class data using *TinkerPlots* software, the task requested students to write a small report predicting characteristics of middle school students, regarding the attributes under study that were the subject of the initial questioning. They were also asked to present evidence to support their conclusions. These reports aimed at pushing them further to see their own data as evidence for making inferences about the whole school, building their ability to think "beyond the data".

Following the initial questions on possible differences between genders regarding the characteristics under study, students compared the two samples (the boys and the girls in the class) and established the possible existence of differences. Afterwards, they generalized this to the school population in this way: "Boys tend to be taller than girls, based on our class data" (WW; Anita & Nelo) and "For middle school students, boys' height and arm span are bigger when compared with girls', based on our class" (WW; Ana & Edu). Students' answers suggest that they considered the data collected in their class as evidence to predict that the population would have similar characteristics regarding height, foot size, and arm span. Further, it is interesting to note that while generalizing, some students already used terms such as "tend to be" in order to express uncertainty, and focused their attention on distribution instead of on specific measures, as was the case when they formulated the initial questions. This suggests that students associated generalization with uncertainty and that the previous work of organising and describing their own data could have supported a shift to inferential reasoning, helping them to see the data as an aggregate and to take into account the data noise. Although the majority of them had determined the average height of students in the class, some of their predictions went beyond that to include a broader range of values and some references to uncertainty. They assumed a statistical perspective of a trend that generally is true but accepts exceptions, as shown in the following claim: "On average, boys are taller but there are also boys with 1.4 [m] who are shorter than girls, while for girls the minimum value is 1.44 [m]" (WW; Willy & Fausto).

However, many students acknowledged that there were differences between the characteristics of their own class' students and those of middle school students regarding height and foot size distributions: "We predict that the average height of the school's students will probably increase due to students' natural growth" (WW; Manel, Sá, & Paulo) and "Although some differences in height may occur, the average height will be bigger, depending on the grade" (WW; Jade & Dino). In these predictions, students seem to consider data noise and the impact of sample's non-representativeness on the accuracy of population estimates. Although students used their knowledge of their familiar context to justify their claims, they focused on older students' characteristics, including those from the 8th and 9th grades. However, they did not bear in mind that 8th grade students had

similar characteristics to those of their own class and, therefore, they would not be contributing for the foreseen height increase.

4.2. USING DATA AS EVIDENCE FOR GENERALIZATIONS

One of the key principles of informal statistical inference is using data as evidence for claims or predictions about a population made from samples. In the first stage, the students used graphical representations built in *TinkerPlots*, in order to explore the collected data of the class and to answer the initial questions or confirm their predictions (on characteristics of the class).

Considering that students were not instructed to build a specific graphic, it was interesting to note that they were able to create various graphical representations and use them properly for representing and interpreting the data in order to get evidence to support their claims. As the most intuitive graphic, dot plots were the most used by the students for both supporting answers to the initial deterministic questions and confirming predictions on class characteristics. For example, in the graphical representations of Figure 2, a group of students properly chose the variables, the graphic and its scale, and also the *TinkerPlots* tool (percentage) to characterize students' heights (altura). On the left, students selected both variables height and foot size (n_Sapato) to create a scatter plot that enabled them to get evidence to answer the initial question: "The tallest student has the largest foot size" (WW; Berto, Bina, & Remi). On the right, the students represented height in two bins and used the percentage tool to support their answer: "The percentage of students with more than 160 cm is 55%" (WW; Berto, Bina, & Remi). However, the students treated data as "classifiers" since they did not regard them as a whole but used only as a portion of data (e.g., the dot corresponding to the tallest student).
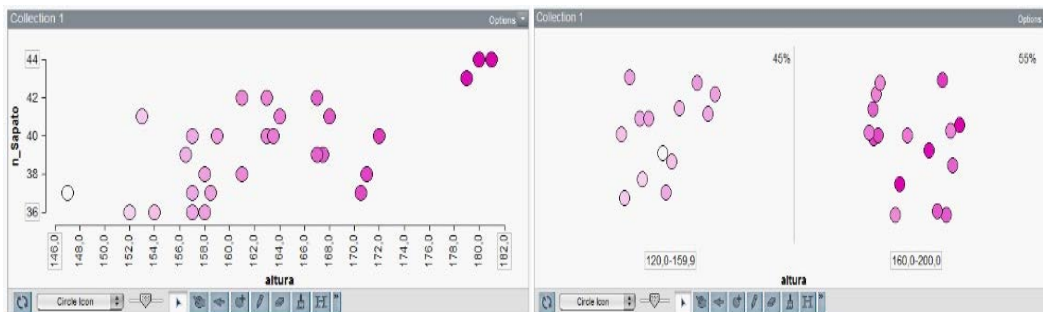

Figure 2. Plots used to characterize students' height (SR; Berto, Bina, & Remi)

Dot plots, like those in Figure 3, were also used by students to compare samples and to confirm their predictions. In the representation on the left, a couple of students claimed that "Based on graphical representations, data shows that it is true that boys are, as a rule, taller than girls" (SR; Anita & Nelo) because they noted that, within their class, the number of girls (F) was lower than the number of boys (M) in the group of biggest heights (160-200 cm, on the horizontal axis). Thus, they confirmed their initial predictions based on direct observation from the data collected in class. However, the number of girls and boys was not the same and they focused on absolute values (number of points) instead of on proportions. Regarding the graphic on the right, which relates foot size (horizontal axis) with height (vertical axis), the same students did a similar interpretation, but this time they considered the proportions. They also made connections

between three variables using the specific software tool of dot colour to distinguish gender, claiming: "Based on graphical representations, data shows that there are more girls than boys with a lower foot size and that the girls also have lower heights, with girls 6 and boys also 6" (WW; Anita & Nelo). They noted that there were twelve students (and explicitly stated that they were six boys and six girls) in the class matching the lower foot sizes, with all six girls in the lower category of height distribution. It is not clear, however, what criteria they used to split the scale into categories.

Figure 3. Dot plots to compare gender regarding height and foot size (SR; Anita & Nelo)

In the graphic of Figure 4, students also used dot plots to compare gender (vertical axis) regarding the arm span (horizontal axis) but included information on arm span averages. Based on the averages, they compared the two distributions, concluding that the arm span is not similar for boys and girls because "boys have an average arm span of 161 cm and girls of 157 cm" (WW; Jade & Dino). In this case, students did not consider data noise, distribution or spread, aspects that could influence the evidence obtained from the data in order to make generalizations.
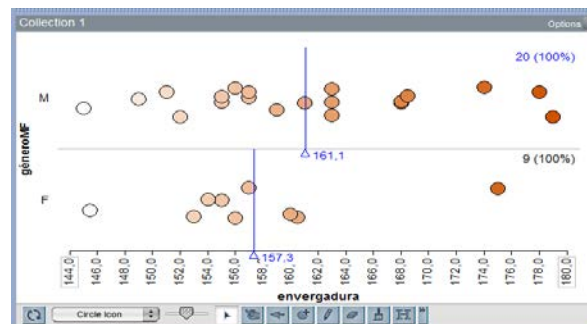

Figure 4. Dot plot including averages to compare gender regarding arm span (SR; Jade & Dino)

Some students built box plots, overlapping dot plots, to compare the two genders regarding different characteristics under study, as presented in Figure 5 for students' heights. However, they did not take advantage of the potential of this type of graphic to compare distributions as they only noted special values such as averages or maximum and minimum values. Other students focused on the maximum and the minimum, concluding that "The girls' minimum height is 154.0 cm and the maximum is 172.0 cm" (WW; Cati & Rudi). Other students focused on the average value, claiming that "Boys

are taller than girls. Boys' average is 163 and girls' average is 160 cm" (WW; Edir & Selma).
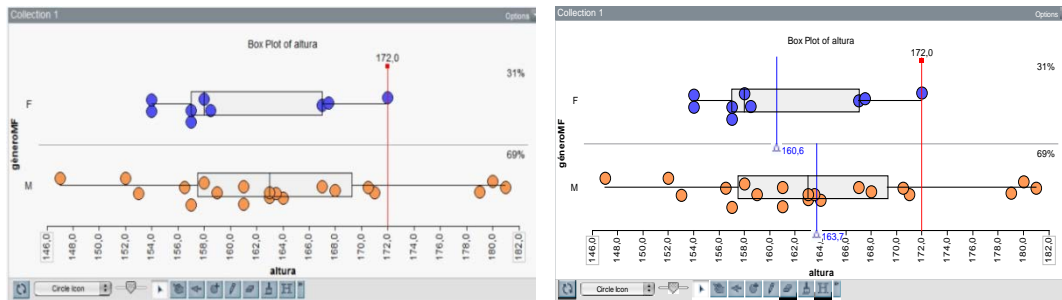

Figure 5. Box plots to compare gender regarding height (SR; Cati & Rudi; Edir & Selma)

These findings show students' difficulties in interpreting box plots and using them in the inferential process, although they had had a previous lesson on building and interpreting box plots. Despite the identified difficulties in this comparison activity, in a phase following the data exploration, students were able to support their claims about the population through analysis of the class's data: "Yes, boys tend to be taller than girls, based on class data" (WW; Tadeu & David) or "Boys are going to be taller than girls, therefore, the average will be on the boys' side because probably there will also be more boys studying than girls. This is based on the fact that in our class there are more boys than girls and that the boys are taller" (WW; Dani, Bia, & Beta). Furthermore, their argument to explain how their claims or predictions were valid was based on data and generalizations, incorporating previous and contextual knowledge. One of the groups argued, albeit incorrectly (as mentioned at the end of the previous section), that "The [population] height average will increase [with respect to the class] because middle school includes 8th and 9th grades, that is, students will be getting taller, having a bigger arm span and a bigger foot size" (WW; Tadeu & David). This was based on their knowledge of 9th grade students' greater heights, connecting it with the observation that data from the class demonstrated a positive association between two of the following variables: arm span, height and foot size. Another group, which showed better understanding of the population under study, considered both older and younger students. Using their knowledge of median as the central value of a distribution, they claimed: "In middle school I think it will be similar since our class is an 8th grade, middle school median. The students are not that different since it is a grade up and a grade down" (WW; Raúl & Ringo).

The exploration of the class' data, guided by the initial questions and encouraged by the visualization of graphical representations built in *TinkerPlots*, supported students' generalizations on the observed characteristics and use of such data as evidence.

4.3. USING PROBABILISTIC LANGUAGE TO DESCRIBE GENERALIZATIONS

Statistical inference also includes using probabilistic language to describe generalizations. Students' reports in this task included this component. Various references to uncertainty were noted in the conclusions drawn about the population under study.

The quotes from students' reports, presented in the previous sections, indicate different ways of dealing with uncertainty inherent to the inferences made from the collected data. Some students used deterministic language when asking questions about the sample or making predictions about the population from the sample. For example, when asking: "Does the biggest foot size match the tallest student in the class?" (WW; Berto, Bina, & Remi), students focused their analysis on individual characteristics and did not feel the need to express uncertainty. Likewise, when inferring that "Boys will be taller comparing to girls based on our class" (WW; Alex & Alfeu), students seemed to believe that the sample gives them full information on population, and did not express uncertainty about the information obtained from it.

The majority of students, however, showed doubts or hesitations about drawing inferences from the sample, acknowledging that their predictions about the population are experimental. They used informal terms like "probably", "maybe" or "similar" instead of equality to express uncertainty. In the following dialogue between two students in the same group we note that they were expecting that their personal experience would be similar to what had been anticipated although accepting the possibility of exceptions (AR; Afonso & Cila):

Afonso: Yes, boys are much taller than girls, usually.
Cila: Usually… I've already seen a boy that was shorter than me, and he was older.

The use of the term "tend to be" in various claims, for example "Boys tend to be taller than girls, based on our class data" (WW; Anita & Nelo), also shows that students were starting to assume a statistical perspective of signal.

Such terms were also the first signals of probabilistic terminology use. We noted, thus, that when the students considered data as an aggregate, identifying a signal in the data distribution, they tended to express their inferences using a word that indicated uncertainty. However, it was not always possible to identify whether those words had an everyday source or were motivated by considering the data noise.

It is also interesting to note the role of context and graphics in students' use of uncertainty expressions. Graphics were the basis of their inferences, encouraging the view of the data as an aggregate and the identification of signal in the data, generalized to the population by the students expressing uncertainty. The following dialogue (AR; Anita & Nelo) shows the conflict between the students' contextual knowledge and the data, leading to the emergence of uncertainty expressions, such as "in general":

Anita: Ok. Based on graphical representations, boys are taller [than girls].
Nelo: Is foot size greater the greater the height? (…)
Anita: Ya! What? This is wrong! He's short and has an enormous foot size!!(…)
Nelo: So, what are we going to say?
Anita: I don't know! This is really confusing and messed up!
Nelo: Let's go: "Based on graphical representations…"
Anita: Slow down, wait a minute. Let's think. 1.60m – 2m… You see!? Big boys, between 1.60m to 2m, have a foot size between 36 to 38, three of them! Edu is there in the middle group…(…)
Anita: Here we find some boys between 1.60m and 1.80m who have big foot sizes. There, a big size. And here two, between 1.60m and 1.80m, have a very little foot size. There. And then, here is another… (…)
Nelo: So, what are we going to say?
Anita: Ah! That's it. "Based on graphical representations, data shows that it is true

that boys are, in general, taller than girls."

On another occasion, the same group acknowledged that their contextual knowledge was confirmed by the data, seen as an aggregate, and argued, expressing uncertainty, that "As girls are on average shorter than boys, they probably have a lower arm span" (WW; Anita & Nelo).

On the other hand, students' inferences seem to be based on the observed evidence of the graphical representations, which encouraged their confidence. The students, using their contextual knowledge, sought deterministic causes for their inferences, thus reducing the level of uncertainty of what data could show regarding the population. This is clear in the claim "The boys have the highest arm span since arm span follows height and boys are taller" (WW; Ema, Carla, & Rui), after observing the graphic relating arm span and height built by the group.

Considering students' age and knowledge, uncertainty is naturally expressed in a qualitative way, with neither confidence levels nor margins of error. Because quantification of uncertainty takes longer to emerge, these informal expressions used by students could be related to an initial perception of a deterministic nature of statistics and could be interpreted as a pioneering version of probabilistic language. Nevertheless, this was the IIR component least shown in the class, which was somehow predictable because students' school path in Mathematics hardly includes a contact with probabilistic ideas or language.

5. DISCUSSION

This study presents the aspects of informal statistical inference that emerged during students' work on a statistical investigation (Wild & Pfannkuch, 1999) using *TinkerPlots* software. Students were engaged with the task and even the ones who showed more difficulties making inferences within the investigation process were capable of displaying some ISI aspects (Makar & Rubin, 2009). This seems significant because these students had little experience in this type of process.

As students progressed through the task, they seemed to understand the value of the inferential process in drawing conclusions about the unknown population (the whole school) based on data collected in class. When selecting the sample, the majority of students acknowledged data noise as a factor to be taken into account in this process, associating it with the uncertainty inherent in the generalization of their results. Thus, students did not seem to have conflicts when drawing conclusions from a very limited sample, their own class, although a further exploration of the confidence level of their inferences is needed. This aspect has been recently discussed in ISI literature (e.g., Manor et al., 2014).

As already noted in other studies (e.g. Watson, 2008), students' initial questions focused on describing the class. This may have resulted from the option of collecting their own data beforehand, diverting students' attention from the inferential process. However, it did not prevent students using data in an inferential way in a later phase of the task, as noted by Watson with younger students. Actually, organising and describing their own data, together with their familiarity of the context and knowledge and mastery of statistical ideas (such as data noise, distribution, spread and graphics, among others), seem to have supported a shift in their attention from specific measures, in order to characterize their class, to seeing data as an aggregate, and taking data noise into account. These latter components are considered essential to inferential reasoning (McPhee & Makar, 2014). The study also shows indications that the exploration of the class' data,

guided by initial questions and supported by visualizations built in *TinkerPlots*, encouraged students to generalize from observed characteristics using the data as evidence rather than base their predictions only on personal experience. The students' acquaintance with this software in two previous tasks seems to have contributed to their meta-representational proficiency (English, 2014), since they tried different representations until they found the one that they regarded as the most useful to support their claims. However, they showed some difficulties in interpreting and utilizing box plots in the inferential process. Such difficulties, also noted in other studies with students from different ages and even with teachers (Pfannkuch, 2006; Watson, 2008), suggest that specific work regarding these representations is needed.

A crucial aspect in this study was the students' articulation of uncertainty when engaging in ISI, which was uncovered by features of the proposed task adding value to the investigation about this topic. It became apparent that asking students to make predictions about the characteristics of the population encouraged them to express uncertainty. The results show that although students' reasoning is ruled by deterministic features, they also express uncertainty in their inferences, using informal probabilistic terms, when they already see the data as an aggregate. It is not clear, however, whether such terms had an everyday source or were motivated by considering the data noise, that is, to what extent they arose from a contextual concern or a statistical uncertainty (Manor et al., 2014). Graphical representations, supported by the use of *TinkerPlots*, also seem to have a key role when students use uncertainty expressions, but do not necessarily promote them. On the one hand, they facilitate students seeing data as an aggregate and help them to express uncertainty in their generalizations as they identify data patterns and trends that are generally true but admit exceptions. In this way, the generated graphs provided students with additional insight about their initial predictions and strengthened their levels of (un)certainty in claims. On the other hand, the power of graphical evidence boosts the confidence of those students who seek deterministic causes for their inferences, using their contextual knowledge. Thus, the use of probabilistic language was the component of ISI least shown by students. Furthermore, the emerging uncertainty in students' inferences was always qualitatively expressed and they did not feel the necessity to quantify it (Manor et al., 2014). Therefore, there is a need to carry out specific work regarding the language of uncertainty, helping students to move away from a deterministic inference perspective and to develop the use of probabilistic terms and appropriate levels of confidence in their claims (Dierdorp et al., 2012).

Lastly, these findings, although limited to one class and one task, provide some understanding of students' capabilities and the challenges they face regarding ISI as they begin to experience Statistics as an investigative process. Additionally, results of this study shed some light on how statistical investigations carried out in a typical classroom setting can effectively encourage students to articulate uncertainty in the context of making ISI, and thus contribute to the recent discussion within the educational community on new ways of articulating probabilistic and statistical ideas.

ACKNOWLEDGEMENTS

REFERENCES

Ainley, J., & Pratt, D. (2014). Expressions of uncertainty when variation is partially-determined. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9A1_AINLEY.pdf

Ben-Zvi, D. (2006). Using *TinkerPlots* to scaffold students' informal inference and argumentation. In A Rossman & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/icots7/2D1_BENZ.pdf

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - Mathematics Education, 44*, 913–925.

Braham, H. M., & Ben-Zvi, D. (2015). Students' articulations of uncertainty in informally exploring sampling distributions. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 57-94). Minneapolis, MN: Catalyst Press.

Cobb P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting design experiments to support teachers' learning: A reflection from the field. *The Journal of the Learning Sciences, 18,* 165-199.

Common Core State Standards Initiative (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers.

Dierdorp, A., Bakker, A., van Maanen, J., & Eijkelhof, H. (2012). Supporting students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. In *Pre-proceedings of the 12th International Congress on Mathematical Education*, Seoul, South Korea. Retrieved from http://www.icme12.org

English, L. D. (2014). Promoting statistical literacy through data modelling in the early school years. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 441-457). Dordrecht: Springer.

Fielding-Wells, J. (2014). Where's your evidence? Challenging young students' equiprobability bias through argumentation. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from
http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2B2_FIELDINGWELLS.pdf

Fitzallen, N., & Watson, J. (2014). Extending the curriculum with *TinkerPlots*: Opportunities for early development of informal inference. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from
http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8D3_FITZALLEN.pdf

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, R., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education: A Pre-K-12 Curriculum Framework*. Alexandria, VA: The American Statistical Association. Retrieved from http://www.amstat.org/education/gaise

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht: Springer.

Henriques, A. C., & Oliveira, H. (2013). Prospective teachers' statistical knowledge for teaching when analyzing classroom episodes. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 41-48). Kiel, Germany: PME.

Kazak, S. (2015). A Bayesian inspired approach to reason about uncertainty: 'How confident are you'? In K. Kraimer & N. Vondrova (Eds), *CERME9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 700-706). Prague, Czech Republic: HAL Archives.

Konold, C., Higgins, T., Russell, S., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305-325.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration*. [Computer software] Emeryville, CA: Key Curriculum Press.

Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Educational Research Journal, 9*(1), 46-67.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1/2), 152-173.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

Manor, H., Ben-Zvi, D., & Aridor, K. (2014). Students' reasoning about uncertainty while making informal statistical inference in an "Integrated Pedagogic Approach". In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8C2_MANOR.pdf

McPhee, D., & Makar, K. (2014). Exposing young children to activities that develop emergent inferential practices in statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, USA). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from
http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2A3_MCPHEE.pdf

Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics*, *88*(3), 385-404. doi: 10.1007/s10649-014-9551-5.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nilsson, P. (2012). Challenges in seeing data as useful evidence in making predictions on the probability of a real-world phenomenon. *Statistics Education Research Journal, 12*(2), 71-83.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman, & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from
http://iase-web.org/documents/papers/icots7/6A2_PFAN.pdf

Prodromou, T. (2013). Students' emerging expressions of uncertainty while making informal statistical inferences about data. In S. Forbes, & B. Phillips (Eds.), *Proceedings of the Joint IASE/IAOS Satellite Conference*: *Statistics education for progress*. Macao, China: IASE/IAOS.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman, & B. Chance (Eds.), *Working cooperatively in statistics education* (Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil). Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/icots7/2D3_RUBI.pdf

Stephan, M., Bowers, J., Cobb, P., & Gravemeijer, K. (2003). *Supporting students' development of measuring conceptions: Analyzing students' learning in social context.* Reston, VA: The National Council of Teachers of Mathematics.

Watson, J. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7*(2), 59-82.

Watson, J., & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*, 44-70.

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

Wisdom Software Inc. (2013). *AutoScreenRecorder 3.1 Pro.* [Computer software] Available at: http://www.wisdom-soft.com/products/autoscreenrecorder_free.htm

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.

ANA HENRIQUES
Instituto de Educação, Universidade de Lisboa
Alameda da Universidade
1649-013 Lisboa
Portugal

APPENDIX

TASK – The human body: a study in school

The *Vitruvian Man* is a famous drawing by Leonardo da Vinci, around 1490, included in his diary. The drawing depicts a man in two superimposed positions with his arms and legs apart and inscribed in a circle and square.

It is based on a famous excerpt of the ancient Roman architect Vitruvius' *Book III* of his treatise *De Architectura,* describing the ideal human proportions. For example, that book specifies that:

- a palm is four fingers
- a foot is four palms
- the length of the outspread arms (arm span) is equal to the height of a man.
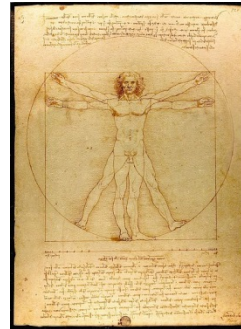
Vitruvius had already tried to fit the human body proportions in a circle and a square but his attempts were not perfect. It was Leonardo da Vinci who correctly adjusted it within the expected mathematics patterns.

(http://pt.wikipedia.org/wiki/Homem_Vitruviano_(desenho_de_Leonardo_da_Vinci))

How could you characterize the middle school students in your school regarding some of Vitruvius' measures such as height, foot size and arm span?
_____

Part I

1. Think about what information will be needed to answer this question and how to collect data by answering the following questions.
   a) What is the population under study?
   b) What sample size can we work with?
   c) How could you choose a representative sample?
   d) What variables should we study? Are these variables qualitative or quantitative? Are they continuous or discrete?

2. Indicate a procedure that would lead to the selection of a biased sample (unrepresentative).

3. For convenience, today we are only going to collect data on our class. Carefully measure and record each attribute (height, arm span, foot size) for each class member.

Part II

1. Analyze the data you collected, which is already in a *TinkerPlots* database.
   a) What interesting questions about this information could you ask? Consider the following:
      - Students' height;
      - Boys' and girls' arm span;
      - Relationship between students' foot size and height;
      - Another aspect that you think is relevant to study.
   b) What do you think is the answer to your questions? Explain the reasons for your answer.
   c) Respond to two of the questions asked in 1a) using graphical representations.

2. From the data collected on your class, prepare a short essay on what you could say about the characteristics of all the students in the school, considering the aspects in question 1a). Explain on what you based your predictions.