

# Chapter 3

## Students Reasoning About Variation in Risk Context



José Antonio Orta Amaro and Ernesto A. Sánchez

**Abstract** This chapter explores students' reasoning about variation when they compare groups and have to interpret spread in terms of risk. In particular, we analyze the responses to two problems administered to 87 ninth-grade students. The first problem consists of losses and winnings in a hypothetical game; the second is about life expectancy of patients after medical treatments. The problems consist of comparing groups of data, and choosing one in which it is more advantageous to bet or to receive medical treatment. In this research we propose three levels of students' reasoning when they interpret variation. Decision making in the third level of reasoning is influenced by risk. As a conclusion, some characteristics of the problems and the solutions provided by the students are highlighted.

**Keywords** Middle school students · Reasoning · Risk · Statistics education  
Variation

### 3.1 Introduction

In this study we analyze the responses provided by middle-school students (ages 14–16) to two elementary problems regarding comparison of data sets. In the problem design, we sought to promote the consideration of spread in comparing two data sets by proposing data sets with equal means and in a risk context. The analysis of the responses consists in identifying and characterizing the students' reasoning levels when they face such problems in order to understand how their reasoning on spread can improve. Spread refers to the statistical variation for data sets and

---

J. A. O. Amaro (✉)

Escuela Nacional para Maestras de Jardines de Niños, Mexico City, Mexico  
e-mail: [jaortaa@gmail.com](mailto:jaortaa@gmail.com)

E. A. Sánchez

Departamento de Matemática Educativa, Centro de Investigación y de Estudios  
Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico  
e-mail: [esanchez@cinvestav.mx](mailto:esanchez@cinvestav.mx)

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, [https://doi.org/10.1007/978-3-030-03472-6\\_3](https://doi.org/10.1007/978-3-030-03472-6_3)

is one of the seven fundamental concepts of statistics (Burrill and Biehler 2011). As Watson (2006, p. 217) indicates, variation is the reason why statistics exists because it is ubiquitous and therefore, present in data. Garfield and Ben-Zvi (2008) observed that “Understanding the ideas of spread or variability of data [...] is essential for making statistical inferences” (p. 203). However, perceiving and understanding variability comprises a wide range of ideas. For instance, there is variability in data, samples, distributions and comparisons of data sets (Ben-Zvi 2004; Ben-Zvi and Garfield 2004; Shaughnessy 2007). This work is focused on the role of spread in the comparison of data sets.

In general, statistics problems comparing data sets involve deciding whether two or more sets can be considered equivalent or not. One way of doing so is through the comparison of center, spread and shape (Ciancetta 2007). Nevertheless, talking about the equivalence of data sets is difficult in basic school problems (Garfield and Ben-Zvi 2005). Therefore, among the problems used in research are those in which the subject is presented with two or more data sets and has to identify the one that has the highest quantity or intensity of the characteristic to be measured (e.g., grades, money, life expectancy). The statistical procedure to solve this type of tasks is based on the calculation of the mean of each set and the subsequent comparison of means. Regardless of its apparent simplicity, finding and reasoned using of this procedure represents a real difficulty for students in basic levels, who are inclined to other strategies (some of which are merely visual while others are based on isolated data from each set) instead of using the mean (Gal et al. 1989; Watson and Moritz 1999). After the study by Watson and Moritz (1999) there has been an increased interest to integrate the role of variation in the students’ analysis when they solve problems about data sets comparison. In this study, we carry on to pose a question with the same objective, proposing new problems to explore the students’ reasoning. We have particularly chosen problems in a risk context to evidence the uncertainty that spread usually uncovers. We then ask: How do students consider data spread in problems involving comparing data sets in a risk context?

## 3.2 Background

Gal et al. (1989) studied the intuitions and strategies of elementary-school students (3rd and 6th grades) when facing tasks of comparing data sets. The tasks were presented in two contexts: outcomes of frog jumping contests and scores on a school test. Several tasks were formulated per context and in each of them, two data sets were presented in a graph similar to those in Fig. 3.1. The students were asked whether both groups performed well or whether one did it better. Characteristics such as number of data, shape, center and spread of each couple of data sets were systematically manipulated to observe their influence on the students’ reasoning. The responses were divided into three categories: *statistical methods*, *proto-statistical methods*, and *other/task-specific methods*. *Statistical methods* included responses in which the sets were compared through data summaries for each set, particularly when arithmetic



R2: All available information, from both visual comparison and calculation of means, was integrated to support a response in comparing groups of unequal sample size. (Watson and Moritz 1999, p. 158)

The differentiation between problems for data sets with the same size and unequal size is related to the use of means to compare. Still, students' consideration of spread was not involved in building the hierarchy. Three years later, Watson (2001) carried out another study, exploring the reasoning of students who had been interviewed in the first research (Watson and Moritz 1999) three years before. In this longitudinal study, besides formulating the research questions of the previous study, Watson posed the following question: "What evidence is shown that variation displayed in the data sets is explicitly considered in making decisions about which group did better?" (p. 343). The students' responses were clustered into six categories: (1) *No acknowledgment of variation*. (2) *Individual features—single columns* [of data]. (3) *Individual features—multiple columns* [of data]. (4) *Global features—'more'* [assumed to be based on visual comparisons]. (5) *Global features—multiple features*. (6) *Global features—integrated, compared and contrasted*. Watson (2002) deepened the same study using a new method, exploring whether the students' responses improve when there is an intervention in which a cognitive conflict in the subjects is created. Each student was shown a video of an interview with another student whose ideas were different from those of the student watching the video, in an attempt to make the student reflect and consider the possibility of changing his or her own conceptions.

Shaughnessy et al. (2004) conducted research to develop and study the conception of variability in middle and high-school students. The students answered a survey, and instructional interventions on sample distribution were carried out; finally, a sub-group of students was interviewed. Particularly, in the second interview, the researchers used the movie waiting time task shown in Fig. 3.2.

The responses to this task were coded in six categories: *Specific Data Points*, *Context*, *Centers*, *Variation*, *Distribution*, and *Informal Inferences*. Such categories were not mutually exclusive, so one response could be coded in two or more categories. When a response was based on the comparison of isolated points of each distribution, it was coded in *Specific Data Points*. When it referred to the student's personal experiences, it was classified in *Context*. Those responses using the means or medians were coded in *Centers* while those including comparisons of variation relative quantities (as considering ranges) were coded in *Variation*. If considerations on center and variation were combined, the response was coded in *Distribution*. Finally, if the response speculated on the probability of having certain experience in the waiting times at each movie theater and the subject used language including terms as 'predictable,' 'consistent,' 'reliable,' 'chances' or 'luck', it was coded in *Informal Inferences*. As a result, they found that most of the interviewed students considered both center and variation in their responses. Two-thirds of the interviewees stated that the two data sets were different from each other despite having the same mean and median. About 70% of them chose the movie theater with the least variation (Royal Theater). Nearly a third of the students included personal experiences in the responses.

**Movie Waiting Time.** A recent trend in movie theaters is to show commercials along with previews before the movie begins. The *wait-time* for a movie is the difference between the advertised start time (like in the paper) and the ACTUAL start time for the movie. A class of 21 students investigates the wait-times at two popular movie theater chains: Maximum Movie Theaters and Royal Movie Theaters. Each student attended two movies, a different movie in each theater, and recorded the wait-times in minutes below.

<p>Maximum Movie Theaters:</p> <p>5.0 12.0 13.0 5.5 9.5 13.0 5.5</p> <p>11.5 8.0 8.5 14.0 13.0 8.5 7.0</p> <p>8.5 12.5 13.5 11.5 9.0 10.0 11.0</p> <p>Mean=10 minutes; Median=10 minutes</p>	<p>Royal Movie Theaters:</p> <p>11.5 11.0 9.0 10.5 8.5 11.0 9.0</p> <p>10.5 9.5 8.5 10.0 11.5 10.5 8.5</p> <p>9.0 11.0 11.0 9.5 10.0 9.0 11.0</p> <p>Mean = 10 minutes; Median = 10 minutes</p>
--	---

**Wait-Time for Movies**

Maximum Movie Theaters

				X						X				
	X			X					X	X				
X	X		X	X	X	X	X	X	X	X	X	X	X	X
5	6	7	8	9	10	11	12	13	14					

Minutes (rounded to the nearest half-minute)

Royal Movie Theaters

				X			X							
				X			X							
			X	X		X	X							
			X	X	X	X	X	X						
5	6	7	8	9	10	11	12	13	14					

Minutes (rounded to the nearest half-minute)

- What can you conclude about the wait-times for the two theaters?
- One student in the class argues that there is really no difference in wait-times for movies in both theaters, since the averages are the same. Do you agree or disagree? Why?
- Which of these theater chains would you choose to see a movie in? Why?

Fig. 3.2 Movie wait time task (Shaughnessy et al. 2004)

Orta and Sánchez (2011) explored how the notions of mean, range and uncertainty influenced the understanding of the statistical variability among middle-school students. The problem with which they collected the information was an adaptation of the Movie Wait Time Task by Shaughnessy et al. (2004). When student participants in the study of Orta and Sánchez (2011) were asked which movie theaters they would choose to watch a film at, and given that the three theaters were at the same distance from home, most of the justifications were based on personal experiences, without taking the data into account. The students’ responses included: “it is nearer home”, “I like those cinemas and I don’t mind watching the trailers they show” and “that

cinema is more famous”. Although the justifications based on personal experiences can be reasonable, in this study they were not based on data.

### 3.3 Conceptual Framework

The statistics education community has distinguished three overlapping areas of statistics to organize and analyze the objectives, activities and results of statistical learning: statistical literacy, reasoning, and thinking (Garfield and Ben-Zvi 2008). This study is located in the area of statistical reasoning. The purpose of the research on statistical reasoning is to understand how people reason with statistical ideas in order to propose features to create learning scenarios. When students try to justify their responses, elements that they think are important to the situation are revealed; in particular, the data they choose, operations made with these data and knowledge and beliefs on which they are based, are important in reasoning.

#### 3.3.1 *Uncertainty and Decision Making*

Statistics is a general method to solve problems in situations where the subject deals with data, variation and uncertainty (Moore 1990). Particularly, Tal (2001) proposes that statistical variation that cannot be explained is uncertainty. We ask: Could considering variation as uncertainty in situations of data set comparison help in decision making? Before answering, we should observe that the contexts from which data arise promote the possibility of associating variation with uncertainty to a greater or lesser degree. For instance, in the context of jumping frogs and scores from Gal et al.’s research (1989), consider problem 1 of Fig. 3.1. It is not easy to think about the variation within each data set as an indicator of the uncertainty. In addition, choosing one or the other option carries no consequences to the person making the decision. Contrastingly, in the movie waiting time task by Shaughnessy et al. (2004), choosing the movie theater whose waiting time data have lesser variation means accepting a lesser uncertainty regarding the start time of the film. Even though the choice, in this case, has consequences for the person making the decisions, the students do not mind the uncertainty and can wait for some time at the movie theater without being affected. They value other characteristics more: closeness of the Movie Theater or comfort of the seats (Orta and Sánchez 2011). To promote the students’ perception of uncertainty and that their consideration will have consequences, it is convenient to choose a context in which the variability of the data involved is more directly related to uncertainty. It is also convenient that choosing one or the other set will have significant consequences. This could be achieved by using situations involving gaining or losing something valuable to the subject, such as health or money.

### 3.3.2 *Tasks in Risk Context*

At first glance, the notion of risk is related to an adverse event that may or may not occur. Aven and Renn (2010) suggest risk is related to an expected value, a distribution of probability as uncertainty or an event. According to Fischhoff and Kadwany (2011), risk is present when there are unwanted potential outcomes that may lead to losses or damages. To make problems involving data set comparison have consequences the subject considers relevant, a risk context seems promising. A paradigmatic task in risk context consists in making a decision about two games where gains and losses are at stake. Consider the following problem:

The gains of realizations of  $n$  times the game A and  $m$  times the game B are: Game A:  $x_1, x_2 \dots; x_n$  Game B:  $y_1, y_2 \dots; y_m$ . Which of the two games would you choose to play?

The solution is reached using a flow diagram: (1) Compare  $\bar{x}$  and  $\bar{y}$ , (2) if  $\bar{x} \neq \bar{y}$ , then choose the game with the greatest mean; (3) if  $\bar{x} = \bar{y}$ , then there are two options: (3a) choose any game, or (3b) analyze the dispersion of data in each game and choose according to risk preferences. Two concepts of the theory of decision on risk (Kahneman and Tversky 1984) characterize risk preferences. Let us say that a preference is motivated by risk aversion when an option with data that have less spread over another with data that have greater spread is preferred. The decision is motivated by risk seeking when the option with data that have greater spread is chosen. For example, in their study Kahneman and Tversky (1984) proposed the participants to choose between 50% of probability of winning \$1000 and a 50% chance of not winning anything and the alternative of getting \$450 with certainty. Many subjects made a decision motivated by risk aversion since they prefer the certain winning, even though the first alternative has a higher expected value.

## 3.4 Method

### 3.4.1 *Participants*

The participants were 87 students (aged 14–16) from two different ninth-grade classes in a private school in Mexico City (last year of middle school). The Mexican middle-school students (7th to 9th grade) study data analysis and graphical representations. They deal with different statistical ideas such as arithmetic mean, range and mean deviation. In addition, they make, read and interpret graphics like bar graphs or histograms (SEP 2011). That is why we expected the students to make use of some of those statistical ideas to explain their answers, and most importantly, to use them in the context (risk). However, the actual responses of students did not meet our expectations.

### 3.4.2 Questionnaire

Two problems (which are presented below) were designed to explore the students' reasoning. Both problems in this research were designed by the authors to study the students' ideas about variation in situations in which "risk or uncertainty" were relevant. The authors considered different contributions of the research in the design. We first used the movie waiting time problem (Shaughnessy et al. 2004) to structure the problems in the questionnaire. When reading and analyzing the movie waiting time problem, we observe it deals with data set comparison. In addition, it has particular characteristics, such as, same number of cases, equal mean and median, and different spreads and bimodalities in both distributions. Ciancetta (2007, p. 103) considers that these qualities were included to promote the reasoning of variation in comparisons. From the structure of this problem (same number of data, equal mean and median, different spreads), we identified these as adequate situations to contextualize the problems of this research.

In the design of the gambling problem, we considered the work by Kahneman and Tversky (1984) in which the authors discuss that an analysis of decision making commonly differentiates between risk and riskless choices. Their paradigmatic example of decisions under risk is the acceptability of a game that leads to monetary results with specific probabilities. In the configuration of this problem, we also considered the idea by Bateman et al. (2007) regarding the introduction of a "small" loss as part of the game. This makes the game seem more attractive and promotes the students' reflection on the situation. Considering the problem of movie waiting time and the gambling situations as reference, we structured Problem 1 of the questionnaire as follows.

**Problem 1.** In a fair, the attendees are invited to participate in one of two games but not in both. In order to know which game to play, John observes, takes note and sorts the results of 10 people playing each game. The cash losses (−) or prizes (+) obtained by the 20 people are shown in the following lists:

Game 1: 15, −21, −4, 50, −2, 11, 13, −25, 16, −4

Game 2: 120, −120, 60, −24, −21, 133, −81, 96, −132, 18

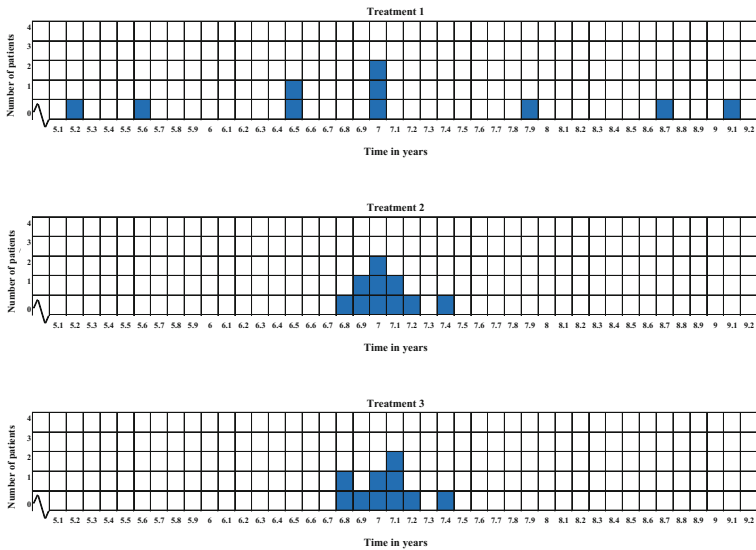
- (a) If you could play only one of the two games, which one would you choose?
- (b) Why?

To create the problem regarding medical treatments, we considered the situation proposed in the research of Eraker and Sox (1981) on scenarios for palliative effects of medication for severe chronic diseases. In this situation, the authors present the choice between drugs that can extend life for several years. With this scenario and the structure of the movie waiting time problem, we created Problem 2 of the questionnaire as shown below.

**Problem 2.** Consider you must advise a person who suffers from a severe, incurable and deathly illness, which may be treated with a drug that may extend the patient's life



Treatment 1:	5.2,	5.6,	6.5,	6.5,	7.0,	7.0,	7.0,	7.8,	8.7,	9.1
Treatment 2:	6.8,	6.9,	6.9,	7.0,	7.0,	7.0,	7.1,	7.1,	7.2,	7.4
Treatment 3:	6.8,	6.8,	6.9,	7.0,	7.0,	7.1,	7.1,	7.1,	7.2,	7.4



- a) Which treatment would you prefer (1, 2 or 3)?
- b) Why?

**Fig. 3.3** Histograms of the three treatments

for several years. It is possible to choose between three different treatments. People show different effects to the medication: while in some cases the drugs have the desired results, in some others the effects may be more favorable or more adverse. The following lists show the number of years ten patients in each treatment have lived after being treated with one of the different options. Each number in the list corresponds to the time in years a patient has survived with the respective treatment. The graphs corresponding to the treatments are shown after (Fig. 3.3).

Table 3.1 shows the statistics of the Problems 1 and 2. Many of the characteristics are the same in both problems.

The general expected solution for each problem is discussed above in the conceptual framework. The problems above were solved in 50 min, approximately.

**Table 3.1** Statistics of Problems 1 and 2

Statistic	Problem 1 games		Problem 2 treatments		
	Game 1	Game 2	Treatment 1	Treatment 2	Treatment 3
Mean	4.9	4.9	7.04	7.04	7.04
Std. dev.	21.43	96.83	1.23	0.17	0.18
Range	75	265	3.9	0.6	0.6
Sum	49	49	70.4	70.4	70.4
Min	-25	-132	5.2	6.8	6.8
Max	50	133	9.1	7.4	7.4
Count	10	10	10	10	10

### 3.4.3 Analysis Procedures

To analyze the responses of students, the chosen group for each of them was first observed; secondly, the responses were categorized in accordance with the strategies of comparison deduced from their justification. For that purpose, we followed the suggestions of Birks and Mills (2011) about identifying important words and groups of words in data for categorizing them and propose levels of reasoning.

The responses were organized to show different levels of the students' reasoning associated with variation. The first shows responses in which variability is not perceived. In addition, few relevant strategies are included to choose one group or another. In the second level, strategies can be considered relevant to choosing between sets of data but ignoring the variability in the data. In the third level, responses show perception of variability and a relevant strategy to choose between one set and the other.

## 3.5 Results

The students' responses were organized in three reasoning levels, considering the type of justification or explanation of the decision or preference made.

*Level 1* groups the responses with circular or idiosyncratic arguments. The first are statements that consider there is a greater gain in the chosen game (Problem 1) or that the treatment allows living longer (Problem 2), but without including data in the argumentation; the second introduces beliefs or personal experiences.

*Level 2* contains responses with justifications that include the explicit considerations of some or all the data in each set. In Problem 1, all the responses at this level obtain the totals and compare them. In contrast, most of the responses to Problem 2 compare isolated data from each set.

*Level 3* is constituted by responses with argumentation combining and comparing more than 1 datum from each set. In these responses, we perceive the differences

**Table 3.2** Responses to Problems 1 and 2 by reasoning level

	Level 1	Level 2	Level 3	No response	Total
Problem 1	64 (74%)	9 (10%)	13 (15%)	1 (1%)	87 (100%)
Problem 2	49 (56%)	25 (29%)	13 (15%)		87 (100%)

between the data sets in terms of the risk each game or treatment involves. Decision making in these cases is influenced by risk aversion or seeking.

Table 3.2 shows a number of responses to the problems in this research that were categorized in each level.

Below, we show examples of the three levels for each problem solved by the students.

### 3.5.1 Problem 1: Levels of Reasoning

**Level 1. Circular argumentation.** In this level, students made a choice but did not justify it based on any treatment of the data. At first glance, the data seemed to suggest that there was more to win or lose (60 responses) in the games. Regardless, they did not specify this and only provided circular arguments such as “because you win more than in 1” (Fig. 3.4). When looking at the data, the students were likely to focus only on some of them and mentally compared them (attention bias). Some characteristics in the responses allowed us to deduce that the students compared specific data. For example: they compared one or two of the greatest data values in a set with one or two from the other. Some focused on the highest losses in each set while others paid attention only to the number of data values with positive (or negative) signs and compared them with the corresponding number of the other set (4 responses).

While in the hierarchy by Watson there are levels that consider numerical and visual strategies, this are surely motivated by the graphic presentation of the data set and detected by the researchers thanks to the interviews. In the case of our study, none of these conditions were presented: the presentation of the data did not include graphs and we conducted no interviews. Visual strategies are those that arise from graphical data observation; however, this research did not include them. However, as stated before, most of the strategies in this level were based on observing, at first

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

El 2  
 ¿Por qué? porque ganas mas que en el 1

**Fig. 3.4** Example of Level 1 response to Problem 1 (“because you win more than in 1”)

Juego 1:		-56								
	15	-21	-4	50	-2	11	13	-25	16	-4 (49)
Juego 2:			-378							
	120	-120	60	-24	-21	133	-81	96	-132	18 (49)

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

cualequiera.  
 ¿Por qué? terminar ganando lo mismo.

Fig. 3.5 Example of Level 2 response to Problem 1 (“you end up winning the same”)

glance, one or two elements in each set and compared each other. In most of the cases, we were unable to determine which specific values students observed in each set because their arguments were circular and there were no interviews afterwards to clarify them. As for the hierarchy by Shaughnessy et al. (2004), Level 1 is similar to the “No acknowledgement of variation” level and partially similar to the “Individual features” level.

**Level 2. Data consideration: Totals.** In this reasoning level, students summed the data in each set and compared the totals (nine responses). From this, students usually argued that it was possible to choose any game because “you end up winning the same” (Fig. 3.5). This strategy enclosed the origin of the statistical procedure of combining observations. For the game problem we analyzed here, the totals are adequate numbers to make the comparison. In addition, we considered all the data and the results were not evident at first glance, but they demanded a certain treatment. The fact that both sets had the same number of data did not allow for differentiating among the students who would perceive the importance of the size of the data sets (proportionality) from those who would not. In the responses placed in this level data variability was ignored or not acknowledged, and risk was not detected.

This reasoning level would be included in Watson’s level M1 because it includes numerical calculations to compare two sets with the same number of data. However, it is not comparable to any level in the hierarchy proposed by Shaughnessy et al. (2004) since mean and median are part of the information given in the data presentation of the problem regarding movie waiting time. Therefore, students did not have to sum data nor obtain the mean, although the reference to mean would appear in some students’ considerations.

**Level 3. Data combination: Risk.** In this reasoning level, we included the responses that provided characteristics to indicate that students perceived risk. In general, the strategies reflected in the responses included in this level consist of simultaneously focusing attention on the relationship between what can be won in each game (maximums) and what can be lost (minimums). The consideration of

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

el juego 2  
 ¿Por qué? porque hay más posibilidad de ganar más, aunque también se pierde más dinero.

**Fig. 3.6** Example of a Level 3 response to Problem 1 (“because there is a bigger possibility of winning more, although you also lose more money”)

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)?

1  
 ¿Por qué? se pueden vivir más años que con otros tratamientos

**Fig. 3.7** Example of Level 1 response to Problem 2 (“one can live for more years than with other treatments”)

choosing game 2 “because there is a bigger possibility of winning more, although you also lose more money” (Fig. 3.6), led the students to notice that the games were not equivalent and risk in one or the other differentiates the games. The choice is not entirely determined by the student’s analysis, but also by their risk preferences. For example, a student chose game 1 as the most convenient and justified the choice by stating “Because as game 2, [game 1] has losses, but in a lower number and you don’t risk that much”. The student’s choice was influenced by *risk aversion*. Interestingly, all the responses, with the exception of one, did not mention the equality of the totals or means of the data sets. This did not affect the analysis that they were equal in both sets. In the only response that considered both the totals and the spread, the student chose any game “because you have the same chance of winning or losing; in one you don’t win or lose much, in the other you win a lot or lose a lot”. In the characteristics written, we perceive that the student summed and obtained 49 in both data sets. Despite noticing that one game was riskier than the other, the student did not prefer any.

### 3.5.2 Problem 2: Reasoning Levels

**Level 1. Circular argumentation.** In this level we included the responses in which a treatment was chosen and the preference was only justified with expressions such as “one lives longer” (32 responses were coded with this argument), or by stating the choice “looking at the graph” (five responses) or providing idiosyncratic justifications (12 responses). In any case, those who said “one can live for more years than with other treatments” (Fig. 3.7) did not clarify why one would live longer with one treatment or with the other. Some other responses did not specify which part of the graph they considered relevant. In the responses coded as idiosyncratic, they introduced personal beliefs that were not relevant to the problem.

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? el 2  
 ¿Por qué? porque ahí me aseguran por lo menos 7 años de vida más

**Fig. 3.8** Example of Level 2 response to Problem 2 (“because you ensure at least 7 years more of life”)

In this problem of medical treatments, the graphical representation of the data promoted the use of visual strategy among the students. However, they did not often explain what they saw in the graph. This level is similar to the level U1 in Watson’s hierarchy, with the difference that she obtained the data through interviews, so that the students could reveal their strategy.

**Level 2. Data consideration.** In this level we classified the responses in which there was argumentation to favor one treatment based on specific values of each set and their comparison (23 responses). We also included the responses in which the data from each set were summed, and the totals were compared (two responses). From the 23 responses above, eight were based on the observation of the maximums and the minimums. For example, a student chose treatment 1 “because there is a greater chance of living for 9 years”. In 15 responses, the modal value was mentioned: a student chose treatment 2 “because you ensure at least 7 years more of life” (Fig. 3.8), but we did not know how the modal value was used to make the decision. We supposed the student combined it with the observation of one or two extremes without stating so.

**Level 3. Data combination: Risk.** In this level, we grouped the responses based on the consideration of two or more data values that allowed students to perceive variability. In 11 responses, the choice was justified mentioning both the maximum and the minimum. For example, a student chose treatment 2 “because I will probably won’t live 9 years more, but I can ensure from 6.8 to 7.4 years.” Two responses mentioned the mode and some extreme; these students chose treatment 3 because the mode was higher than in the other two, but they also considered some extreme values. As an example, a student chose treatment 3 “because in the graphs the minimum is 6.8 years and the most frequent [period] is 7.1 while in treatment 1, 5.2 is the minimum and 7 is the most frequent.” In several of these responses, the students provided characteristics that allowed us to suppose that they perceived risk; for example; when they stated “because I may not live 9 years more, but I still have from 6.8 to 7.4 for sure” (Fig. 3.9) and chose treatments 2 and 3, avoiding the risk of living only 5.2 years, as with treatment 1. We can suppose that their choice was influenced by risk aversion. On the other hand, the responses in which they chose treatment 1, and that considered that they “can live up to 9 years although [one] could live only for 5.2 years”, were influenced by risk seeking.

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? Tratamiento 03  
 ¿Por qué? Porque tal vez no viva 09 años más  
pero así tengo asegurada de 6.8 a 7.4

**Fig. 3.9** Example of Level 3 response to Problem 2 (“because I may not live 9 years more, but I still have from 6.8 to 7.4 for sure”)

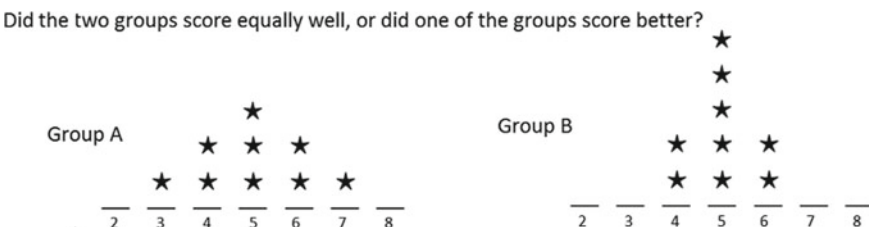
### 3.6 Discussion

It is convenient to make some observations that may be useful to answer the question posed at the beginning of this work and clarify the purpose of this research. Particularly, we will consider some characteristics of the problem and the solutions provided by the students.

In relation with the characteristics of the problem, we emphasize that spread can be significantly interpreted to use it in decision making in risk contexts. In the context of winning in games, we easily perceive that the game with the highest spread is riskier because there is more to win, but there is also more to lose. In contrast, in the context of score problems by Gal et al. (1989) and Watson and Moritz (1999), it is unclear how the spread of data can allow making a decision on the group that performs better.

In data set comparison problems, when the sets to compare have the same number of data and same mean (and median) but have different spreads, it is unclear how spread is used to make a decision. For instance, Gal’s problem (1989) presents the data in Fig. 3.10; these data represent the score that students in Groups A and B obtained in an exam. The question posed is: Which group performed better?

In Group A, a student scored 7, better than any other in Group B; however, another student obtained a score of 3, which is worse than the score of any other student in Group B. There are three students in Group A whose score is the mean while there are five students in Group B in the same condition. Do these observations have implications that allow deciding which group is better? They apparently do not clearly help to decide if one group performs better than the other when no other criteria are added. If the consideration on variation has no implications, then why should it be done?



**Fig. 3.10** One of the problems of Gal et al. (1989) adapted

We have observed that, when the data context is winning a game or life expectancy after a medical treatment, several students place themselves in position where they are benefitted or affected by one decision or the other. They realize that the game or treatment corresponding to the set with the highest spread offers them the possibility of winning more or living longer, but also of losing more or living for a shorter time. On the other hand, the game or treatment with the least spread data set offers them the least uncertainty. As in Gal's problem, there is no regulatory outcome to indicate which the best decision is because such response depends on the risk preference of the person making the decision. This is an uncertainty pattern frequently present in daily life: Do we prefer a stable job or business, even if we earn a little or one in which we earn more but carries greater risk? The concepts of risk aversion or seeking help us to understand how we decide when facing those options.

The reasoning levels that have been defined are characterized by the presence or absence of data in the justification as well as their combination. Given that what we analyze are the students' written responses not their thoughts, the levels reflect their ability to write the idea that leads them to make a decision. As mentioned before, the hierarchy we propose, is different from the ones by Watson and Moritz (1999) and Shaughnessy et al. (2004) in that they conducted interviews, hence the nature of their data is different.

In *Level 1*, the students' argumentation does not include the data they used to make the decision. Most of the students "justify" their decisions using circular arguments, that is, arguments that only claim what they should prove. This shows that many of them are not conscious that a key point of argumentation is to evidence of the way in which data were used to make a decision. They surely saw something in the data that affected their response, but they did not write that in their response.

In *Level 2*, the choice and the use of data to reach a conclusion are shown in the argument. However, either the selection is not the best to make a decision or the use of data is not the best. In this level we grouped the responses with arguments that do not consider variation. A number of them sum the data in each set and compare them. This strategy is partial because it allows the students to perceive the equality of the means because the sets have the same number of elements, but the students fail to consider the spread. Others only consider isolated data, as comparing the maximums, the minimums or the ranges. If they considered more than one of these data, they did not link them properly; therefore, they did not perceive spread.

*Level 3* shows the choice and use of data in the argument to reach a conclusion. The data are also combined to provide a notion of spread and risk. The students did not use a measure of spread but coordinated more than one data value, mainly maximums and minimums, to perceive that the games in situation 1 and the treatments in situation 2 were different due to the risk they posed. In this way, the measure of spread they implicitly used was the range and with it they perceived the risk.

In the design of the present study we sought to reveal the students' reasoning regarding spread in data set comparison, and we did not try to reveal their reasoning about and with the notion of center, particularly the arithmetic mean. Regardless, the total absence of a strategy to calculate the arithmetic mean of each data set and the comparison between them is remarkable. The closest thing to this strategy is



calculating the total of each set and comparing them. Clearly, students do not think of these totals as representatives of each set but as the amount of money or time involved in them. Although technically, the only missing element is dividing between the total data to have a comparison based on the means, the conceptual distance between the idea behind the consideration of the totals and the idea of representative of a data set is quite large.

### 3.7 Conclusion, Limitations and Implications

Based on the results obtained in this research, the risk context provides an alternative with which students engage and in which the words or numbers used by them to solve the problems are numbers and words in context, a central idea in statistics. When solving the problems raised, the students associated risk and variability; the latter was described through the range of the data sets. As part of the analysis, the students used the sum of all the data available; however, they did not manage to complete the algorithm for the arithmetic mean. There is a void in the students' analyses; they do not use center and dispersion of a data set for making a decision. Research deriving from this should pose a learning trajectory that guides students to integrate the notions of center and dispersion. This would enable students to improve their analysis of data sets or distributions.

The presence of students' reasoning considering risk shows the convenience of looking for situations in this context to suggest problems related to dispersion. Together with other scenarios already used in teaching to promote learning and consideration of dispersion, the concept of risk might be another source of problem situations for statistics teaching at mid-school level. Accordingly, a deeper research might find other risk situations to explore students' reasoning and promote their development in the statistics class. Additionally, the reasoning levels students can reach when facing those problems and situations could be characterized with greater precision. Additionally, the reasoning levels students can reach when facing those problems and situations could be characterized with greater precision. The use of technology is a necessary complement for a more thorough research that allows for easier calculations and emphasis on conceptual discussions. Furthermore, technology provides options for simulations, which could correspond to distributions related to the game or medical treatment contexts.

The risk models that truly work in society are very complex, even though gains/losses situations are common in business. Decisions and consequences hardly derive only from such situations despite the fact that the calculation of probabilities, centers, and dispersion are frequently part of the analysis. This is a limitation in developing the option of promoting risk situations in statistics teaching at middle school. Still, we consider its possibilities have not been sufficiently explored for statistics teaching at other educational levels.

**Acknowledgements** Grant EDU2016-74848-P (FEDER, AEI). Grant CONACYT No. 254301.

## References

- Aven, T., & Renn, O. (2010). *Risk, governance and society*. Berlin: Springer.
- Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making*, 20(4), 365–380.
- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63.
- Ben-Zvi, D., & Garfield, J. (2004). Research on reasoning about variability. *Statistics Education Research Journal*, 3(2), 4–6.
- Biggs, J. B., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behaviour. In H. Rowe (Ed.), *Intelligence, reconceptualization and measurement* (pp. 57–76). Hillsdale, NJ: Laurence Erlbaum Associates.
- Birks, M., & Mills, J. (2011). *Grounded theory*. Thousand Oaks, CA: Sage.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 57–69). New York: Springer.
- Ciancetta, M. (2007). *Statistics students reasoning when comparing distributions of data* (Doctoral thesis). Portland State University. Online: <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/07.Ciancetta.Dissertation.pdf>.
- Eraker, S. E., & Sox, H. C. (1981). Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making*, 1(1), 29–39.
- Fischhoff, B., & Kadavy, J. (2011). *Risk: A very short introduction*. Oxford: Oxford University Press.
- Gal, I., Rothschild, K., & Wagner, D. A. (1989). *Which group is better? The development of statistical reasoning in elementary school children*. Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO.
- Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92–99.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341–350.
- Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
- Orta, A., & Sánchez, E. (2011). Influential aspects in middle school students' understanding of statistics variation. In M. Pytlak, T. Rowland & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society For Research in Mathematics Education, Rzeszów, Poland*.
- SEP. (2011). *Programas de Estudio. Educación Básica. Secundaria. Matemáticas*. México, D. F.: Secretaría de Educación Pública.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1010). Greenwich, CT: Information Age Publishing, Inc. and National Council of Teachers of Mathematics.
- Shaughnessy, J. M., Ciancetta, M., Best, K., & Canada, D. (2004). *Students' attention to variability when comparing distributions*. Paper Presented at the Research Pre-session of the 82nd Annual Meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.
- Tal, J. (2001). *Reading between the numbers: Statistical thinking in everyday life*. New York: McGraw-Hill.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47(3), 337–372.

- Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, 51(3), 225–256.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.
- Watson, J.M., & Moritz, J.B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168.