FACTORS CONSIDERED BY SECONDARY STUDENTS WHEN JUDGING THE VALIDITY OF A GIVEN STATISTICAL GENERALIZATION

Hanan Innabi

ABSTRACT. . This study investigated the factors that 12th grade students in the United Arab Emirates take into consideration when judging the validity of a given statistical generalization, particularly, in terms of the sample size and sample selection bias. The sample consisted of 360 students who had not studied sampling yet. Results show that a small percentage of the students take the sample size and selection bias into consideration properly. Many students based their judgment on their personal beliefs regardless of the properties of the selected sample. This study identified some pre-teaching misconceptions that students have with regard to sampling. Such misconceptions include 'any sample represents the population', and, 'any sample does not represent the population'.

KEYWORDS. Statistical Thinking, Sampling, Judging Generalizations, Secondary School Students.

BACKGROUND

The realization of the importance of Statistics as a subject that helps students use data and information to communicate and make decisions led the educational system in the United Arab Emirates (UAE) to give more attention to teaching statistics and probability at schools. Accordingly, statistics education in the UAE is going through a crucial change stage. Educational reform related to Mathematics and Science curricula and textbooks has been going on since the year 2000 in the UAE. Old mathematics textbooks are gradually being replaced by new ones set by the Ministry of Education. It is expected that in the coming few years, new textbooks will be enforced for all the 12 grades.

At all levels, statistics and probability are part of the mathematics curriculum. A comparative look at statistics and probability in the old and new mathematics curricula shows a big difference. Teaching and learning statistics according to the old curriculum are carried out in the traditional way. Statistics for both teachers and students focuses on calculations and rules to be applied. There is no use of *real data*, and the use of the computer in teaching or learning statistics does not exist in the schools. According to the old curriculum, students in the UAE start

learning statistics from 6th grade and probability from 11th grade. The attention given to statistics and probability chapters in each grade is usually less than given to the other mathematics chapters. The statistical content that the students study according to the old curriculum consists of separate unrelated concepts and generalizations, in addition to some applications and drills using theoretical, pre-prepared and unreal data. The content in the old curriculum is primarily descriptive statistics, including *basic graphics, frequency tables, central tendency measurements,* and *spread measurements*.

The new mathematics curriculum promises a better picture. According to the new curriculum, students in the UAE are studying statistics and probability from the first grade instead of the 6th grade. Many of the statistical ideas that were not included at all in the old curriculum are present in the new curriculum. In the new document of mathematics curriculum outlines, one can find a new spirit related to learning statistics and probability. In this document and within the 'Data analysis and probability standard' one can find sub- standards and performance indicators that relate to statistical thinking, such as formulating questions, collecting, organizing, representing, analyzing, and interpreting data, assessing statistical inferences and predictions, and understanding and applying basic concepts of probability (MOEU, 2001). This promising picture is a consequence of *Data Analysis and Probability* being one of the ten standards upon which the new mathematics school curriculum in the UAE is built. One of the ideas that was not included in the old curriculum and will be presented in the new curriculum for secondary school students is *sampling techniques* and *standard error*. These concepts are basic to understand inferential statistics.

This study is based on the idea that students have previous interests, ideas, conceptions and misconceptions about *sampling*. Teaching statistics has to help students recover, enhance or restructure this previous knowledge. This study provides some information about how secondary students think about the relationship between sample and population and their conceptions and misconceptions related to sampling techniques. It is hoped that this information will be useful for the ongoing mathematics reform, particularly for mathematics teachers training programs and curriculum developers.

THE PROBLEM

For both descriptive and inferential statistics, the terms *sample* and *population* are the initial and basic terms. Understanding of these terms and of the nature of the relationship between them is an important condition for students to understand many of the statistical concepts and procedures (Lovett & Greenhouse, 2000). The relationship between the sample and population depends on the idea that the sample, as a part of the population, can be examined in order to obtain a generalization true of the population, or what is called a *statistical generalization*.

As in all inductive inferences, we cannot establish that a statistical generalization is true with absolute certainty. Our concern, usually, is about how likely it is that the conclusion is valid. The crucial feature that determines the strength of a statistical generalization is the representativeness of the sample. In another words, to what extent are the features of the population that concern us reflected accurately in features of the sample (Lewis, 1999; Salmon, 2002). Usually, it is not easy to tell whether a sample is representative. However, two criteria are considered noteworthy: 1) the sample is large enough; 2) the sample is varied enough. In some cases, a very small sample can support a strong generalization; in others, a very large sample is required. The real question is whether the sample is large enough to capture, or represent, the variability present in the population.

The use of statistical generalizations appears not only in scientific research, but also in daily social issues, where many conclusions are based on a sample of behaviors or observations (Nisbett & Ross, 1980). Accordingly, one can consider statistical generalization as an aspect of critical thinking. Ennis (1985) in his taxonomy of critical thinking considered the *Inducing and judging induction/ Generalizations/ Sampling* one of the main abilities of critical thinking.

Thus, it is important to direct efforts towards helping students improve their reasoning related to *statistical generalization*. The improvement of such reasoning demands an adequate understanding of the mechanism by which the student reasons, the variables that affect it, the teaching methods that enhance it and evaluation techniques that measure it.

This study tries to provide some knowledge of how secondary students think when they judge the validity of a given statistical generalization. In particular, this study tries to answer the following question: What factors do secondary grade students in the UAE, who have not studied sampling methods, take into consideration when judging the validity of a given statistical generalization?

LITERATURE REVIEW

Living in an uncertain world requires the ability to reason statistically. Many pervasive and persistent errors are found in people's reasoning about uncertainty (Konold, 1991; Seldmeier & Gigerenzer, 1997; Innabi, 1999). Research has shown the weaknesses in people's reasoning in social situations which demand statistical reasoning (Kahneman & Tversky, 1972, 1973, 1974, 1979; Nisbett & Ross, 1980; Konold, 1989; Kuhn, 1991).

Evidence is provided to show that statistical training helps to improve the use of statistical principles in reasoning. Research showed that training is useful whether it consisted of several statistics courses or a single course or even a short training session. Also training in a given domain can transfer fully to another domain. Research also showed that the training not only enhanced statistical thinking for college students but also for high school students and adults (Shaughnessy, 1981; Lehman & Nisbett, 1990; Nisbett, 1993; Fong, Krantz & Nisbett, 1993; Lawson et al, 2003; Rubin, Hammerman, & Konold, 2006).

Reviewing the literature related to people's judgments in some statistical situations show that one or more of the following factors have been investigated; sample size, sample variation, sampling error, and personal experiences. Kahneman and Tversky presented a series of experiments on intuitive statistical judgment in which subjects did indeed ignore sample size. They found that for statistically naïve college students, the similarity of a sample statistic to a population parameter does not depend on the size of the sample. They explained people's reasoning under uncertainty by two heuristics; *representativeness* (probabilities are evaluated by the degree to which an object A belongs to class B, that is, by the degree to which A resembles B), and *availability* (assessing the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind) (Kahneman & Tversky, 1973, 1979; Tversky & Kahneman, 1973, 1974).

The view that statistically naïve people ignore sample size has been modified by complementary research. A number of studies have shown that subjects may take account of sample size if the form of the problem is modified or when the variable is manipulated in alternative tasks (Olson, 1976; Evans & Dusoir, 1977; Bar-Hillel, 1979; Nisbett et al, 1993).

Cosmides and Toody (1996) showed that people can perform better in their judgment under uncertainty simply by expressing the problem in frequentist terms (frequentists arguing that probabilities refer to the long-run relative frequencies of events in the world whereas Bayesians argue that probabilities refer to subjective degree of confidence). From the Frequentists' point of view, humans may be good intuitive statisticians.

Well, Pollatsek, and Boyce (1990) conducted a series of experiments. Different versions of the problems were presented to undergraduate students who had not previously taken a college statistics course. The results suggested that naïve subjects' appreciation for the law of large numbers often does not result from in-depth understanding of the relation between sample size and variability.

An early indication of people's insensitivity to considerations of randomness versus bias in sample selection came from a study by Nisbett and Borgida (1975). Nisbett and Ross (1980) reported that the sample bias is an even more important concern than the sample size. They maintained that people fail to apply necessary statistical principles to a very wide range of social judgments. They claimed that people often make overconfident judgments about others based on small and unreliable amounts of information; they are often insensitive to the possibility that their samples may be highly biased.

Rubin, Bruce, and Tenny (1990) explored some of the underlying conceptions and heuristics students bring to the study of statistics. They organized their investigations around a set of concepts about sampling that are basic to understanding statistical inference. The central idea of statistical inference that they depended on is that a sample gives us *some* information about a population- *not nothing* and *not everything*. They investigated students' naïve conceptions of sampling representativeness and variability. The results showed that students

have inconsistent models of the relationship between samples and populations. Their answers in different problem settings fall in varying amounts under the influence of intuitions about sample representativeness or sample variability.

In addition to the above research that considered the sample size and the sample variation as factors related to some statistical reasoning, there is some research that is concerned with personal experiences. This research showed that personal perspective and personal narrow experiences led individuals to be biased in their judgments (Evens, 1989; Falk, 1998). Shaughnessy (1992) pointed out that the individuals do not perceive events that happen to them as just one more tally in a big objective frequency distribution.

The present study investigates the accurate and inaccurate conceptions that students bring to the topic of sampling before they are instructed in this topic. This study was designed in light of knowledge gained from the previous research described above. Considering that the above previous knowledge came mainly from psychological perspectives, one can say that the present study is an educational implementation that can be useful for teaching and learning statistics at the school levels.

METHOD

Definitions and variables

The following definitions and variables are used in this research:

Statistical generalization: A statement made about the population from the knowledge obtained from the sample.

Sample size: Number of the elements in the sample. Two types of sample size have been considered in this research; a relatively small sample with 6 elements and a large enough sample with 600 elements.

Sample bias: The tendency for a sample to differ from the population due to the sampling design –that is, structure, size, and method of selection of sample elements. In this research the term 'bias' refers to 'selection bias'. Two types of samples have been considered in this research; biased sample (sample from the students who were entering the library where the population of interest is all the university students) and unbiased sample based on random selection.

Instrument

To answer the research question an instrument was designed to explore the factors which influence students when judging the validity of statistical generalizations. The instrument contained a written problem that presented information about a sample and the relevant population. Then a conclusion about the population was presented based on a sample statistic.

This problem was written in a context of a real problem from life in order to capture the effect of personal factors as previous research has shown this may affect people's reasoning.

As described in the literature review, the previous research explored two factors that may affect people's statistical judgments: sample size and sample bias. To investigate whether the students consider these two factors, the problems in this study were constructed to provide information about sample size and bias. Thus the problem has alternative versions, each with different information about the given sample.

It has been noticed that the previous research that studied people's statistical reasoning asked the subjects to compare between two samples with different sizes and distributions. In this study it was important not to direct students to any specific factors as the purpose is to capture these factors. Accordingly, the problem was built to contain only one sample and a conclusion and the students have to be asked to judge the validity of this conclusion and give all the justifications.

Based on the above considerations, the instrument contained only a written problem. The context of this problem was about a student in the University of UAE who was interested in the number of visits that students make to the University library during the term time. Therefore he selected a sample of University students and asked each of them *how many times do you visit the University library every week during term time*. The average number of weekly visits in the problem sample was three, so the student concluded that the average number of visits the UAE University students make to the library during term was approximately three visits per week (assuming that the information given by the sample was reliable). Figure 1 displays an example of these versions.

Instructions. The following problem contains specific information followed by a conclusion that depends on the given information. What you have to do is to judge the validity of the given conclusion, if it is valid or not valid and then write the reasons for your answers.

Please read the question carefully, then select your answer by putting the sign X. Then write all the reasons that made you choose your answer. Remember that it is very important to write all the reasons and not just some of them.

Issue. A graduate research student in the University of UAE was interested in the number of visits that students make to the University library during the term time. Therefore he selected a sample of the University students and asked them 'How many times do you visit the University library every week during term time?' The average number of weekly visits in this sample was three.

> He concluded that the average number of visits UAE University students make to the library during term is approximately three visits per week (consider that the information that was given by the sample's students is reliable).

Judge the validity of the conclusion:

a. Valid conclusion b. Not valid conclusion c. I can not judge

Give <u>all</u> your reasons for this answer:

Figure 1. Instruments' Problem – No information about the Sample version

Since this study was to investigate the factors which students take into consideration when they judge the validity of a given statistical generalization, in particular sample size and the sample bias, the above problem was presented to different students in five alternative forms. The only difference among these versions is the nature of the sample which was selected in each one. These samples are:

• Large/biased sample: The student in the problem took 600 students randomly selected from those who were entering the library entrance.

• Large/not biased sample: The student in the problem took 600 male and female students randomly selected from different scientific and humanistic colleges in different years of study.

• Small/biased sample: The student in the problem took 6 students randomly selected from those who were entering the library entrance.

• Small/not biased sample: The student in the problem took 6 male and female students randomly selected from different scientific and humanities colleges in different years of study.

• No information about the sample size or sample bias: The student in the problem selected a sample of the University students.

The response expected from students was to judge the validity of the generalization as *valid* conclusion, *not valid* conclusion or *cannot judge*. Students were also requested to offer all the reasons justifying their selection.

For versions one, three, and four, the expected answer to the closed question is 'the conclusion is not valid'. For version two the expected answer is 'the conclusion is valid'. For version five where there is no information about the selected sample the expected answer is 'I can not judge the conclusion'. It was expected that students provided a proper statistical explanation for their judgments on the given conclusion such as 'the conclusion is not valid because the sample is small' or 'the conclusion is valid because the sample is big enough and varied'.

For the purposes of validity and reliability of the instrument, three experts in educational psychology who were interested in "reasoning" issues were asked to evaluate the problem with regard to the research question. Their comments indicated that the instrument was suitable. Also, the instrument was re-administered to the same 15 students (these students were not from the original sample). Each student gets just one form. The students' answers for the two times were compared. The result of this comparison showed that the answers in the first time were consistent with the second time in 98% of the answers (students' judgments and explanations).

Procedures

The instrument was given to 360 students from 12th grade (17-18 years old) in the science stream from 12 secondary schools (6 females schools and 6 males schools) chosen

randomly in the city of Alain in the United Arab Emirates. It was important to pick students from different achievement levels. Two school supervisors in Alain city judged that each school in the schools sample represents a typical secondary school that contains all three achievement levels both in mathematics and in general achievement. As each school contained more than one 12th grade class (sections), one section was randomly selected from each school. The size of each section ranged between 28 and 33 students. The sample students had studied the basic concepts in descriptive statistics and basic probability concepts in an unrelated way. They had not studied any content related to sampling techniques, sampling error or sampling distributions. The reason for selecting these schools was that the concepts of sampling techniques and sampling errors will be part of the new curriculum in the secondary school. Providing information about the previous perceptions, understanding, and misunderstanding that the secondary students have regarding the relationship between sample and population will help in developing the curricula and the teacher training programs.

Seventy-two copies of each form of the test (5 forms) were prepared. To be sure that each form will be presented equally, 6 copies of each form were prepared for each of the schools and were distributed randomly to the students of the selected section. Data was collected in January 2005. A day before the test administration; the students were informed by the school educational counselor that they would participate in research that aimed to improve the Mathematics curriculum. A researcher would give them a written test. To be sure that the test atmosphere and instructions were the same in all the classes, the researcher administered the test in all the 12 schools.

The average time needed for the test was 20 minutes, 10 minutes for the instructions and 10 minutes for answering the problem. In the first 10 minutes, the researcher explained the importance of the students' help to get valid results which may be useful to improve mathematics teaching. After distributing the test, the written instructions on the first page were read through with the students.

By using the blackboard, the nature of the problems and what the students had to do was clarified. The focus was placed on the importance of putting *all* the reasons for their judgment. It was noticed that the students in general were interested and serious while answering the problem.

Data Analysis

Two stages of analysis were carried out. In the first stage, each student's explanation (answer) was coded into four codes as follows:

1. The sample size code: (0 or 1). If the student's answer took into consideration the size of the sample as a factor to judge the validity of the conclusion, she/he was given the code 1, otherwise 0.

2. The sample bias code: (0 or 1). If the students' answer took into consideration the sample bias as a factor to judge the validity of the conclusion she/he was given the code 1, otherwise 0.

3. The other factor codes (1 to *n*): Any factors other than sample size and bias that the students presented in their explanations were written down and given a code from 1 to *n*. Since only a few of the students put more than two *other* reasons for their judgment, it was decided to open two columns (variables) for the 'other factors' data.

To test the coding reliability, the researcher recoded 20 cases which were randomly selected. The comparisons between the two codes indicated a 100% consistency. The data were entered to the computer using the SPSS packages. The actual number of cases became 338 after some improper cases were deleted. The paper was judged as improper case if it was empty or contained completely irrelevant sentences and words.

To get a clearer view of the students' answers and to summarize the data, the second stage of the analysis was carried out. It was necessary to look at the answers as a whole within each of the five forms (large\bias, large\no bias, small\bias, small\no bias, no information about the sample) and according to each judgment (valid, not valid, cannot judge). Students' answers were coded again by one or more of the following categories that are clarified in the results section:

1. Adequate statistical explanation (code 1).

2. Insufficient statistical explanation (code 2).

3. Personal belief explanation (code 3).

4. Inadequate statistical explanation (code 4).

5. No clear answer or no explanation (code 5).

It is important to note that in the first stage of the analysis the student's answer was coded regardless whether it was correct or incorrect. In this stage the coding process was done without considering the form or the judgment. The purpose in this stage was to determine (by coding) the students written explanations or in other words the factors that students considered. In stage two of the analysis, new codes were generated by looking into six variables; forms, judgment, sample size, sample bias, other factor 1, and other factor 2.

RESULTS

Stage One

The percentages of students who mentioned in their written explanations the sample size factor or the sample bias factor (took the size or bias factors into consideration) are given in Table 1.

Table 1. Percentage of students who considered sample size and bias

Type of Sample	N. of students taking this item	Percentage of students	
		Consider Size	Consider Bias
Large/biased	70	14	7
Large/not biased	69	22	9
Small/biased	68	59	18
Small/not biased	66	48	8
No information	65	28	12
Total	338	34	12

From Table 1 we see that the 'size' factor is considered more often than the 'bias' factor. In general, around one third of the students took the size factor into consideration, while 12% of the students took the sample bias into consideration. When the sample size was small, the percentages of students who took this factor into consideration were higher than when the sample size was large.

Stage Two

Table 1 gives the percentages of students who took the sample size and bias into consideration, but it does not provides a clear picture about how students used these two factors to judge the validity of a given generalization. For example, the students who considered the sample size factor in the *large/biased sample* and said that the conclusion was *valid* because the sample was large enough, could not see that this sample cannot represent the population even though the sample size is large. A deeper analysis is required in order to understand more about students reasoning in such situations. This analysis led to a categorization of students' explanations into five categories; *Adequate statistical explanation, Insufficient statistical explanation, Personal belief explanation, Inadequate statistical explanation, and No explanation*.

Adequate Statistical Explanation: A proper explanation to support a student's judgment on the given conclusion was categorized under adequate statistical explanation. - i.e. when the student gave the *expected* answer in the closed question supported by a proper statistical explanation using the sample size and/or bias factors, their explanation was considered as *adequate statistical explanation*.

Specifically, in the case of the large and unbiased sample, answers which judged the conclusion as *valid* because the sample size was adequate and the sample selection was unbiased were considered as adequate statistical explanation. In the case of the large and biased sample, answers that judged the conclusion as not <u>valid</u> because the sample was biased were considered

as adequate explanation. In the case of the small and unbiased sample, answers, which judged the conclusion as not *valid* because the sample was small, were considered as 'adequate explanation'. In the case of the small and biased sample, explanations that mentioned at least one factor (small or biased) to explain why the conclusion was not valid were also considered as adequate statistical explanation. In the case of 'no information about the sample' the students who mentioned that they could not judge the conclusion because the way the sample was selected was not clear or the sample size or the potential for bias was not addressed were considered also as adequate explanation.

Insufficient Statistical Explanation. Some explanations that students provided were not enough to support their judgment; such an explanation was considered as an *insufficient statistical* explanation. Insufficient explanations were found in two situations: Large biased and Small unbiased samples.

In the case of the large and biased sample the response which mentioned that the conclusion was *valid* because the sample was large was considered an insufficient answer. The response *I cannot judge* because the sample is both large and biased (with the meaning that there were two factors, one of them proper and the other not) was considered also as an insufficient answer. In the case of the small and unbiased sample the response which mentioned that the conclusion was *valid* because the sample was unbiased was also considered as an insufficient statistical answer. The response which mentioned that *I cannot judge* because the sample is small but unbiased was also considered as an *insufficient statistical* answer.

Personal Belief Explanation. Many students provided explanations that reflected their personal opinion or experience about the subject of attending the libraries. For example the following explanations were considered as *personal* explanations: the conclusion was valid because: "the number of visits is reasonable /expected, there are negative attitudes towards the library", "youth do not worry about reading", "students need the library", "the library is useful", "I go to the library" or the conclusion was not valid because: "I have a brother/sister/friend in the University who has never entered the library", "I think that the average of the number of visits should be more (or less) than three", or I cannot judge the conclusion because: "I am not in the University", "I do not know the university student's need for library", "I have no idea".

Inadequate Statistical Explanation. This category contained the responses which used statistical explanation incorrectly. In the following we present some examples of inadequate statements that the students provided to explain their judgments:

• The conclusion is valid because: "Any sample of the students represents the whole students in the University", "any part represents the whole", "He selected from the students who were entering the library", "his selection from those who go to the library makes the conclusion stronger", "he took the average, because he took more than one opinion".

- The conclusion is not valid because: "He should take all the students in the University", "any part does not represent the whole", "if he had selected any other 600 students he would have found another result", "He should select a much bigger sample", "the actual number of the visits may be more or less".

- I cannot judge because: "Any part represents the whole", "He selected from the students who were entering the library", "He took the average", "He took more than one opinion", "He should take all the students in the University", "any part does not represent the whole", "if he had selected any other 600 students he would have found another result.

The percentage of students who considered that *any sample would not represent the population* was 18% and the percentage of students who considered the opposite (i.e. *any sample will represent the population*) was 4%.

Table 2 presents the numbers of students according to their explanations. One can notice that in some cases students' explanations about their judgments contained more than one reason, each from a different category. For example in the situation of the large and biased sample some students said that the conclusion was valid because the sample was large and also because they believed three visits was a reasonable number. Such a response carried insufficient explanation (valid because the sample is large) in addition to a personal explanation (they themselves think the average number of visits in the conclusion is a reasonable one).

As can be seen from Table 2 in the case of the large/ biased sample just 4 students out of 70 were able to justify properly why the conclusion was not valid. Notice that 39 students judged correctly the conclusion as not valid. Nevertheless most of them (33) justified their judgment in an inadequate or personal way. In the large/ unbiased sample around half of the students who gave the expected answer did not provide adequate explanations for their answer. In the case small/ unbiased sample, around half of the students who answered in the expected way did not present an *adequate* explanation. In the case of small/biased sample more than one third of the students who answered as expected that the conclusion is not valid, did not provide an *adequate* explanation. Around three quarters of the students who answered as expected in the *no information* about the sample did not provide an *adequate* explanation.

Table 2. Numbers of the students according to their explanations' categories

Type of sample in the problem	Judgment	Adequate	Inadequate	Insufficient	Personal	Adequate and Personal	Insufficient and Personal	Inadequate and Personal	No explanation	Total
							Type of Explanation			
Large / biased	Valid	0	1	7	6	0	2	0	1	17
	Not valid	4	19	0	11	1	1	1	2	39
	Can't judge	0	5	0	4	1	1	1	2	14
Small / biased	Valid	0	5	0	2	0	0	1	0	8
	Not valid	33	11	0	4	2	0	3	1	54
	Can't judge	0	3	0	1	0	0	0	2	6
Large/ unbiased	Valid	6	1	0	3	0	0	0	1	11
	Not valid	0	21	0	3	0	0	5	2	31
	Can't judge	0	14	0	4	0	0	0	1	19
Small/ unbiased	Valid	0	2	8	1	0	1	0	1	13
	Not valid	21	13	0	3	2	0	0	2	41
	Can't judge	0	10	1	6	0	0	0	3	20
No information	Valid	0	4	0	3	0	0	0	2	9
	Not valid	0	27	0	9	0	0	1	2	39
	Can't judge	5	5	0	3	1	0	1	2	17
Total		69	141	16	63	7	5	13	24	338
		20%	42%	5%	19%	2%	1%	4%	7%	100%

In general it seems from Table 2 that one fifth of the students presented an adequate statistical explanation for their judgment, while two fifths of the students presented an inadequate statistical explanation and around one quarter of the students presented a personal explanation.

DISCUSSION

This study investigated the factors that the secondary students in the UAE take into consideration when judging the validity of a given statistical generalization. Results showed that students considered some factors related to sample size, sample bias, inadequate factors (like; any sample represents the population or any sample does not represent the population), and personal experiences and expectations.

Sample Size and Sample Bias

Results showed that the percentage of students who mentioned the sample size factor was 34%. Assuming that students wrote all of the reasons that led them to their judgments, it can be said that two-thirds of the students did not see the sample size as a factor that affects the validity of the statistical generalizations. A similar statement can be made about the sample *bias* factor, as only 11% of the students correctly took the sample *bias* into consideration. These results support the idea that sample characteristics are apparently not part of a person's repertoire of intuitive ideas (Kahneman and Tversky, 1972 and Evans, 1989).

When information about sample size and bias was given, it was noticed that more students took sample size into consideration in the situations where the sample size was small than in the situations when sample size was large. The percentages of students who took the *size* and *bias* into consideration when no information about the sample was given were 28% and 12% respectively and when the sample was *small* and there was *bias* these percentages were 59% and 18%. In the situation where the sample was *large* and *unbiased*, the percentages were 22% and 9%. These results are consistent with previous research (Bar-Hillel, 1976; Olson, 1976; Evans & Dusoir, 1977; Bar-Hillel, 1982; Well et.al., 1990; Cosmides & Toody, 1996), which showed that the form of the provided problem (framing of problem instructions) affects whether naïve subjects take the sample size into account in their judgments and predictions.

The analysis of students' responses showed that not all the students who took the sample size and bias into consideration provided an adequate justification. Just one fifth of the students considered both factors in an adequate way in forming their judgment. The analysis of students' explanations revealed the following two misconceptions related to the sample *size* and sample *bias*:

1. Some students did not realize that a sample which was clearly biased did not represent the population. They looked at the bias in the *biased sample* as a factor that made the conclusion valid. This observation seems to agree with what Kahneman and Tversky found regarding the representativeness heuristic.

2. Some students used information that was not sufficient to support their judgment. Some of them took only one of the sample properties (size or bias) into consideration in supporting their judgment of the validity of the conclusion and forgot about the other. For example, when the sample consisted of 600 students selected at random from those who were entering the library entrance, many students judged the conclusion as *valid* because the sample size was large without recognizing the *bias* factor in the sample selected.

Inadequate Factors

Two inadequate factors were clear in the students' explanations. Some students (4%) considered that any sample irrespective of its size and bias was a good representation of the

population. These students could not see the differences (variability) among students and could not see that the way the sample was selected made any difference in representing the population. It seems that those students believed that any part could represent the whole without any understanding of the difference between the sample statistic and the population parameter and how the sample properties can affect this difference.

Some students (18%) insisted that any sample would not represent the population. This may indicate that those students could see that the sample *statistic* differs from the population *parameter*, but they could not see that this difference can be reduced through manipulation of sample properties. Students in this category were able to see sampling variability to the extent of leading them not to *believe* in sampling. In other words they have the belief that no *true* knowledge about any population can be obtained through sampling.

We can relate the above two points with what Rubin, Bruce, and Tenny (1990) called sampling representativeness (sample gives everything) and sampling variability (sample does not give anything). This research supports the existence of these two patterns of reasoning among students. However there was a difference between the findings of this research and the research of Rubin, et. al.. In this research, more students said that any sample will not represent the population (sample variability) than said that 'any sample will represent the population' (sampling representativeness), while in the Rubin, Bruce, and Tenny study there was no clear pattern of difference between sampling variability and sampling representativeness intuition.

Personal Factor

The analysis of students' answers showed that the students' personal opinions affected their judgment. The students often used their own experience and expectations to judge the validity of the conclusion. It appeared that students' personal expectations about the population studied affected their judgment so that if the given conclusion matched their expectation, the conclusion was judged as *valid*, otherwise it was judged as *not valid*.

The results showed that 26% of the students provided a personal explanation for their judgment; 19% of the students provided just personal explanations and 7% provided 'adequate or inadequate or insufficient explanation' in addition to the personal explanation.

Part of the effect of personal beliefs on statistical judgment can be explained by the *availability* heuristic suggested by Kahneman and Tversky. Some students, in responding to the problem, used easily accessed information to form their opinions.

The results showed that, in addition to the *personal* explanations, some students (7%) provided other *statistical* explanations (adequate or inadequate or insufficient explanations). Often, these students provided one of these explanations to support the other one. In other words, these students provided *personal* explanations in order to support their *statistical* explanations or

vice versa (i.e., provided statistical explanations to support their personal beliefs). Further study is needed here to fully understand this behavior.

IMPLICATIONS

This research disclosed some misconceptions that students have before starting formal study of sampling techniques at school. Among the misconceptions identified are: any sample can represent the population; no sample regardless of its size and lack of selection bias can represent the population; a conclusion is valid (or not valid) because personal experience supports it; the larger the sample the more valid the conclusion regardless of selection bias. It is hoped that revealing these misconceptions will be helpful to those who write textbooks and to those who teach these topics.

One technique that could be used to change students' misconceptions is to confront students with examples and situations that lead them to see their misconceptions and motivate them to change them. For example when students are taught the definition that a sample is a subset of the population, they also should be given examples of samples contained in the population which are not representative of the population.

An approach that can be followed to probe students' errors is provided by Shaughnessy (1993). He suggests that teachers should include examples of misuses and abuses of statistics in their classes on probability and statistics, and encourage their students to rebut them with correct analysis. He suggests using the problems that have been used in the research as tools to probe students' statistical reasoning errors. In the light of Shaughnessy's suggestions, one application of this research is the possibility of using problems similar to the problems that have been used in the instrument of this research in the classroom to focus students' attention on errors being made in formulating judgments and to clarify how beliefs and conceptions can affect decisions under uncertainty.

It is hoped that the teaching of *sampling* will help students to believe in sampling as a scientific technique that helps people make conclusions about a population, to understand that any conclusion based on sample results involves a degree of uncertainty, and to realize that the validity of a statistical generalization is dependent on the properties of both the sample and the population (see for example, Watson, 2000; Lawson, Schwiers, Doellman, Grady, & Kelnhofer, 2003; Phung, 2005; Reading, & Reid, 2006; Saldanha, & Thompson, 2006). Teachers do not want students to be only able to define the terms *sample* and *population* and to calculate how many samples of size n can be obtained from a population with size N, and to describe methods of sampling without real understanding and without being able to reason critically when encountering a statistical generalization in a newspaper.

REFERENCES

Bar-Hillel, M. (1979). The role of sample size in sample evaluation. *Organizational Behavior and Human Performance*, 24, 245-257.

Bar-Hillel, M. (1982). Studies of representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp 69-83). Cambridge: Cambridge University Press.

Cosmides, L., & Toody, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1-73.

Ennis, R. (1985). A logical basis for measuring critical thinking. *Educational Leadership*, 43 (2), 44-48.

Evans, J. (1989). *Bias in human reasoning: causes and consequences*. London: Lawrence Erlbaum.

Evans, J., & Dusoir, A. (1977). Proportionality and sample size as factors in statistical judgments. *Acta Psychologica*, 41, 129-137.

Falk, R. (1989). Judgment of coincidences: Mine versus yours. *American Journal of Psychology*, 102, 477-493.

Fong, G., Krantz, D., & Nisbett, R. (1993). The effects of statistical training on thinking about everyday problems. In R. Nisbett (Ed.), *Rules for reasoning* (pp 91- 139). London: Lawrence Erlbaum

Innabi, H. (1999). *Students' understanding of the relationship between sample and population*. Unpublished Ph.D. University of Nottingham, England.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive* Psychology, 3, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Kahneman, D., Tversky, A. (1979). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp 201-208). Cambridge: Cambridge University Press.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical Constructivism in Mathematics education* (pp. 139-156). Dordrecht: Kluwer.

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Lawson, T. J., Schwiers, M., Doellman, M., Grady, G., & Kelnhofer, R. (2003). Enhancing students' ability to use statistical reasoning with everyday problems. *Teaching of Psychology*, 30 (2), 107-110.

Lehman, D., & Nisbett, R. (1990). A Longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, 26 (6), 952-960.

Lewis, C. P. (1999). Understanding the sampling distribution and the central limit theorem. Paper presented at the *Annual Meeting of the Southwest Educational Research Association,* San Antonio, TX.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *American Statistician,* 54 (3), 196-206.

MOEU. (2001). *National document on mathematics curriculum.* United Arab Emirates: Ministry of Education and Youth.

Nisbett, R. (1993). Reasoning, abstraction, and the prejudices of 20th century psychology. In R. Nisbett (Ed.), *Rules For Reasoning* (pp. 1-12). London: Lawrence Erlbaum.

Nisbett, R., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32, 932-943.

Nisbett, R., Krantz, D., & Jepson, C. (1993). The use of statistical heuristics in everyday inductive reasoning. In R. Nisbett (Ed.), *Rules for reasoning* (pp 15- 54). London: Lawrence Erlbaum

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Prentice-Hall, Inc, Englewood cliffs, NJ.

Olson, C. (1976). Some apparent violations of the representativeness heuristic in human judgement. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 599-608.

Phung, N. (2005). Public opinion polls, chicken soup and sample size. *Teaching Statistics*, 27 (3), 89-92.

Reading, C., & Reid, J. (2006). Listen to the students: Understanding and supporting students' reasoning about variation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: http://www.stat.auckland.ac.nz/~iase.

Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314-319). Voorburg: International Statistical Institute.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: http://www.stat.auckland.ac.nz/~iase.

Saldanha, L., & Thompson, P. (2002). Students' scheme-based conceptions of sampling and its relationship to statistical inference. *Proceedings of the Annual meetings of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1305-1316) Athens, GA: PME-NA

Saldanha, L., & Thompson, P. (2006). Investigating statistical unusualness in the context of a resampling activity: Students exploring connections between sampling distributions and statistical inference. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: *http://www.stat.auckland.ac.nz/~iase*.

Salmon, M. (2002). *Introduction to logic and critical thinking*, 4th ed. Belmont, CA: Wadsworth.

Shaughnessy, J. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. Schulte (Ed.), *Teaching Statistics and Probability. Yearbook of the National Council of Teachers of Mathematics (NCTM)*. (pp. 90-100). Reston, VA: NCTM.

Shaughnessy, J. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: National Council of Teachers of Mathematics and MacMillan.

Shaughnessy, J. (1993). Probability and statistics. *The Mathematics Teacher*, 86 (3), 244-248.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp 3-22). Cambridge: Cambridge University Press.

Watson, J. (2000). Preservice mathematics teachers' understanding of sampling: intuition or mathematics. *Mathematics Teacher Education and Development,* 2, 121-135.

Well, A., Pollatsek, A., Boyce, S. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes*, 47, 289-312.

Author   : **Hanan Ayoub Innabi**

Address   : Mathematics Education, Department of Curriculum and Instruction,

                     Faculty of Education, United Arab Emirates University,

                     Maqam Campus, POB 17551

                     Al Ain, Abu-Dhabi, United Arab Emirates

Phone   : (++971) 713-4811