ASSESSING STATISTICAL LITERACY: WHAT DO FRESHMEN KNOW?

Eugenia Koleza, Aristoula Kontogianni

Department of Education, University of Patras, Greece

While there are many studies about the statistical literacy of students those who concern the statistical literacy of pre-service teachers are relatively few. With the present study we attempt to investigate the level of statistical literacy of pre-service teachers in their first year at the university and after the end of schooling. For the purpose of this study we adapted the framework for statistical literacy of Watson (1997, 2003) and Gal (2002), while for the assessment of the participants' responses we used a modified model of SOLO taxonomy. Our study findings indicate the low level of statistical literacy among pre-service teachers in their first year at the university.

Keywords: Statistical literacy, pre-service teachers, statistical knowledge, freshmen.

INTRODUCTION

Many researchers (e.g. Watson & Callingham, 2003; Budgett & Pfankuch, 2007) in the field of statistics education have emphasized the importance of statistical literacy for the effective participation of students in the society after the end of school. As the definition of statistical literacy is still being refined (Rumsey, 2002), in this paper we refer to it as the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions (Wallman, 1993). Statistical literacy is a key ability expected of citizens in information-laden societies, and is often touted as an expected outcome of schooling and as a necessary component of adults' numeracy and literacy (Gal, 2002).

Several studies have shown that teachers' knowledge is connected to what and how students learn and depends on the context in which it is used (Ball & Bass, 2000; Cobb, 2000). Consequently, it is important to inquire into the cognitive level of pre-service teachers in order to modify accordingly the content of the courses, which are relevant to Mathematics, at tertiary education.

While there are many studies about the statistical literacy level of students (e.g. Watson & Callingham, 2003) there are relevant few about pre-service teachers. Focusing on the dimension *"as an expected outcome of schooling"* of statistical literacy we designed and conducted a research whose main goal was to identify at what level the freshmen at the Department of Education were statistically literate after the end of the schooling and their entrance to the university. In particular our research questions were:

How able are pre-service teachers to understand fundamental statistical concepts and use them in order to perceive and to criticize information about the world around them? In other words, at what level pre-service teachers are statistical literate?

BRIEF LITERATURE REVIEW

Statistical literacy is a major goal of several curriculum of mathematics around the world (e.g. NCTM Principles and Standards, 2000; ACARA, 2010). According to Australian Curriculum (ACARA 2010, p.2):

> Students should develop an increasingly sophisticated ability to critically evaluate chance and data concepts and make reasoned judgments and decisions. They should develop an increasingly sophisticated ability to critically evaluate statistical information and build intuitions about data.

Following this international trend several studies have been conducted in order to define statistical literacy (e.g. Watson, 1997; Watson & Callingham, 2003) and investigate students' statistical literacy at different levels of education (e.g. Budgett & Pfankuch, 2007 for college students). Furthermore the ARTIST Web site (https://app.gen.umn.edu/artist/) created by DelMas and his colleagues (for more details DelMas et al., 2007) provides and evaluates tools for the assessment of students' statistical literacy. In the field of adults' statistical literacy, research by Gal (e.g. 2002) was a major contribution to the conceptualization of statistical literacy, while Moreno (2002) focused on the connection of statistical literacy with citizenship, and Shield (2006) through the W. M. Keck Statistical Literacy Project, immersed statistical literacy in society. Recently, Kaplan & Thorpe (2010) applied for adults the framework of statistical literacy, proposed by Watson & Callingham (2003).

THEORETICAL FRAMEWORK

As "the research in statistical literacy has unveiled a very deep construct involving a myriad of types and skills and cognitive processes" (Shaugnessy 2007, p.966) for the present study we restricted to a framework based on a combination of the work of Watson (1997) and Gal (2002) in relation to the kind of statistical knowledge students should have by the end of schooling.

Watson (1997) proposed a three-tiered Statistical Literacy Hierarchy:

> **1.** Understanding of basic statistical terminology. **2.** Understanding of statistical language and concepts when they are embedded in the context of wider social discussion. **3.** Ability to question claims that are made in context without proper statistical justification.

Respectively, Gal (2002) suggests that, for full participation in the society, students after the end of schooling should be able to:

(a) … interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant.

(b) … to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions. (Gal 2002, p. 2-3).

For our research we used a framework drawn on the combination of the above theories in order to define our assessment goals and construct the respective tasks. A parameter we took also in consideration was the statistical content of elementary mathematics curriculum that pre-service teachers will have to implement in the future. More precisely, we focused on: the average, the reading and the interpretation of tables and graphs and the critical questioning of claims that originate to social context.

METHOD

In order to answer the research question we designed and conducted a research project during the first semester of the academic year 2011-2012.

Participants

The participants were 166 students (pre-service teachers), 137 female and 29 male at their first year of their studies in the Department of Education. The students were taught the basic concepts of Statistics at the 4^{th}, 5^{th} and 6^{th} grade (ages 10-12) of primary education, the 2^{nd} and 3^{rd} grade (ages 14-15) of Junior High School and the 3^{rd} grade (age 18) of High School. Statistics is taught at the 3^{rd} grade of High School as a part of the course "General Mathematics" which is taught for two hours weekly and it is obligatory for all students regardless of their programs of study (Theoretical, Practical and Technological Direction) (Ghinis et al. 2009). The participants of our research were 134 (80.7%) of Theoretical Direction and 32(19.2%) of Practical/Technological.

In their last year of High School (3^{rd} grade), in the chapter of Statistics, students are taught how to process statistical data and interpret critically statistical conclusions. The Syllabus includes the following subjects (Pedagogical Institute of Greece, 2007):

Basic concepts: The students are taught basic statistical concepts such as population, variables (quantitative and qualitative), census and sample.

Presentation of Statistical data: The students are taught about frequency distributions and their graphical representations.

Location measures and measures of variation: The students are taught how to compute the arithmetic mean, the median, the mode (location measures) and the range, the variance, the standard deviation and the coefficient of variation (measures of variation) of discrete and continuous variables.

Questionnaire

The questionnaire items that we used for this study were either adapted from items used in previous researches (Aoyama, 2003; Watson, & Callingham 2003; PISA contest, 2003; DelMas et al., 2007) or formed by the researchers for the needs of the present study. The questionnaire included ten items, open-ended and multiple-choice. Students (pre-service teachers) were requested to justify their answer for all items. Time given for response was about 1.5 hour. For the coding of the responses we adapted the SOLO model of Biggs & Collins (1982) in the way that Watson & Moritz (2000) have used it. The complexity levels are described in the next table:

Code	Level	Description
4	**Relational**	**Correct justification.**
3	**Multistructural**	**Correct answer with partial justification.**
2	**Unistructural**	**Not able to interpret correctly the data or irrelevant use of data.**
1	**Prestructural**	**No justification. Justification based on irrelevant data or personal estimation.**
0		**No response or Yes/No answer without justification.**

Table 1: Codes and description

Each item had a scoring rubric which was designed to identify increasing quality of response and these varied from 0-2 to 0-4, depending on the complexity of the item. The coding was done independently by two raters. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters. It was found to be Kappa = 0.736 ($p < 0.001$), which indicates substantial agreement between the two raters (Landis & Koch, 1977).

RESULTS

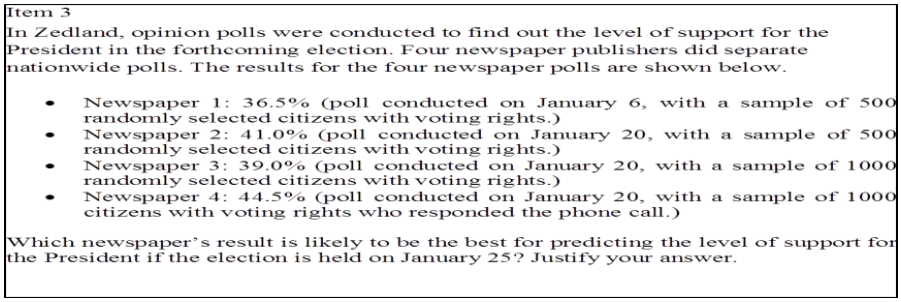
Students' performance was mixed: on four (out of nine items) students' performance was rather low, while on the other five the majority of students' responses were at the relational level. Due to limited space, we analyze only the low- performances' items, while for the high-performances' ones we give only a concise description.

The first two items (adapted from Watson, 2003) concerned the notion of average. In the first item most of the students, 83%, gave answers at the relational level as they were able to reverse the algorithm for the computation of the mean. In the second item 65% of the students gave answers at the relational level, as they were able to choose the mean as the most accurate measure. In the fifth item concerning the interpretation of a line-plot (2 questions) and the computation of the range (1

question) (adapted from Watson & Callingham, 2003) the percentage of answers in the relational level were 97%, 87% and 62% respectively. The sixth item was about the interpretation of a bar-graph (3 questions) (derived from Aoyama, 2003). Most of the students were classified to the relational level: 83%, 72%, and 70% for each part of the item. For the eighth item (derived from Watson & Callingham, 2003) 83% of students gave justifications in the relational level, as they were able to identify that the percentages given in the pie chart were incorrect.

Discussion on items with low performance

The next items, involve contexts that may appear in the media, and their objective was to assess students' "ability to question claims that are made in context without proper statistical justification" (Watson, 1997).



Item 3

In Zedland, opinion polls were conducted to find out the level of support for the President in the forthcoming election. Four newspaper publishers did separate nationwide polls. The results for the four newspaper polls are shown below.

- Newspaper 1: 36.5% (poll conducted on January 6, with a sample of 500 randomly selected citizens with voting rights.)
- Newspaper 2: 41.0% (poll conducted on January 20, with a sample of 500 randomly selected citizens with voting rights.)
- Newspaper 3: 39.0% (poll conducted on January 20, with a sample of 1000 randomly selected citizens with voting rights.)
- Newspaper 4: 44.5% (poll conducted on January 20, with a sample of 1000 citizens with voting rights who responded the phone call.)

Which newspaper's result is likely to be the best for predicting the level of support for the President if the election is held on January 25? Justify your answer.

Figure 1: Item 3

Code	Description	Examples
4	Choice of the third newspaper and correct justification based to the date and to the sample. Discriminating that the sampling that was based to phone calls isn't so accurate.	"The third newspaper because the date is closer to the elections' date and the sample is bigger. It is better than the fourth because the sample method is more reliable (examples why is more reliable)''
3	Correct answer with partial justification. (Referring only to the date or to the sample)	"The third newspaper because the sample is bigger"
2	Using arithmetic data in order to find how many people will vote for the president. Irrelevant justification.	By multiplying percentages with the number of the sample.
1	Choice without reasoning. Choice of the fourth newspaper considering the way of collecting data as more accurate.	'' The fourth newspaper because it is more reliable the sampling method''
0	No response.	

Table 2: Codes and examples to the codes

The third item (derived from PISA, 2003) concerned reasoning about samples. Only the 40% of the participants' answers were classified at the relational level (code 4). The analysis of the students' responses revealed that they had difficulty to understand why a sample cannot be representative of the population.

> Student 65: The fourth newspaper because they used phone calls in order to gather the data.

Also, several students confused the poll's percentage with the number of the sample.

> Student 18: Fourth newspaper because 44,5 % of the voters is almost 500 people more than the others' newspapers population.

The following information is from a survey about smoking and lung disease among 250 people

	Lung disease	No lung disease	Total
Smoking	90	60	**150**
No smoking	60	40	**100**
Total	150	100	**250**

Using this information, do you think that for this sample of people lung disease depended on smoking? Justify your answer.

Figure 2: Item 4

Code	Description	Examples
4	Critically examines all information (implies look at all cells) and/or correctly states proportions and percentages	"No, because smokers and not smokers have the same percentage (90/150=60/100=60%)"
3	Correct answer with partial justification (citing evidence from the data focusing on 2 or 3 cells).	"Yes, because smokers that have lung disease are more than no-smokers".
2	Not able to interpret correctly the data or irrelevant use of data (Referring to the limitations of the methods used to collect the data).	"We can't jump to conclusions relying in such a small sample".
1	Positive or negative statements justified with knowledge of the content area, but not based on data.	"Yes, if you smoke probably you will be sick with lung disease"
0	No response.	

Table 3: Codes and examples to the codes

The fourth item was adapted from the study of Watson & Callingham (2003). It concerns proportional reasoning and only 22% of the freshmen performed at the relational level. Several participants showed difficulties to use correct the given numbers.

> Student 88: In my opinion, no. 100 people of the 250 do not have lung disease and only 90 people are smokers.

While others answered depending to their own opinion instead of use the given data of the table.

> Student 17: The lung disease does not depend on smoking because there are also non-smokers that they have lung disease.

The performance of the students to this specific item was an alarming indication, given that 1) participants to the study were pre-service teachers that will be called to teach proportionality in primary school and 2) proportional reasoning is a necessary prerequisite for performing at the highest level of statistical literacy understanding (Watson & Callingham, 2003).

Item 7

The state based its decision on the redundancy on the data table below, according to which 50 to 66 civil servants who belong to the age group 62-64 want to go out on redundancy voluntarily. Do you agree with this conclusion? Justify your answer.

Reasons for retirement and retirement age

	Retirement Age		
Reasons for retirement	Under 62	Between 62 and 64	65 or more
Age	10,5	21,6	64,6
Redundancy	10,5	50	14,6
Health problems	26,3	11,9	8,3
Close business	10,5	1,5	0
Privileges	10,5	3	0
Opportunity for the younger	2,6	1,5	6,0
Unhealthy labor	5,3	4,5	0
Family matters	7,9	0	2,9
In order to enjoy life	7,9	1,5	2,1
Other reason	7,9	4,5	2,1
	100%	100%	100%
Total	76	66	48

Figure 3: Item 7

Code	Description	Examples
4	Ability to discriminate and use correct the data of the table. Correct use of percentages.	"The conclusion is wrong because the 50% of 66 is 33".
3	Their reasoning is based on the complementary information. They realize that "the rest percentages" don't add to 50, but they cannot extend this reasoning to the rest 50%.	"If you add all the other reasons it is less than 50"
2	Not able to interpret correctly the table.	"I agree because it refers to 50 people in 66".
1	Choice without reasoning. No justification or justification based on irrelevant data.	"It is wrong, nobody would choose redundancy".
0	No response. Yes or No answer without justification.	

Table 4: Codes and examples to the codes

While this specific item was rather easy only the 2% of the participants responded to the relational level. The majority of the participants'answers (43%) were at the prestructural level (code 1). Students justified their judgment by resorting only to their own beliefs (Budgett & Pfannkuch, 2007) rather than giving data-based justifications.

Student 11: It's a lie; nobody wants to go to redundancy.

Student 21: It is impossible so many people to be in redundancy.

The interpretations of this item indicates that, since Statistics is a social construct (Best, 2001), students' reaction is connected to their knowledge of real-life.

Item 10

Two newspapers published articles about criminality and they used the next graphs. Which of these two do you think that represents better the reality? Justify your claim.

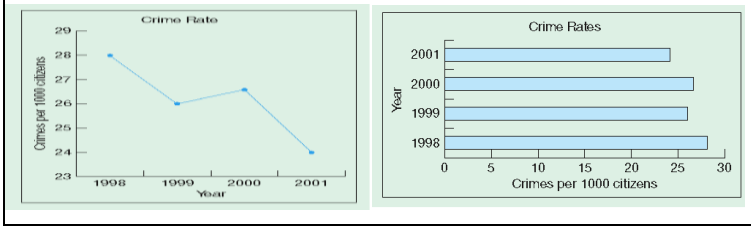


Figure 4: Item 10

Code	Description	Examples
4	Correct justification: ability to discriminate the wrong choice of the scale and the false affect that it creates.	"The bar-graph is better because the scale starts form 0". "The time-graph is not correct, it is like an enlargement"
3	Correct answer with partial justification.	"The bar-graph is better because it is more accurate"
2	Wrong justification based on personal estimations.	"The time-plot, because it is easier to read it, to understand it"
1	Response without justification	" The time plot is better"
0	No response.	

Table 5: Codes and examples to the codes

This item refers to a misleading graph and its aim was to assess the ability of the students to discriminate the false effect that it creates. Only 1% of the participants

were able to answer at the relational level while the majority of them (59%) had chosen the time-plot as they considered it as easier for interpretation.

> Student 45: I believe that the 2nd is better because through it, it is easier to understand the decline of crime rates during these years.

It is a fact that graph interpretation cannot be effective, if the reader does not possess basic graph reading skills. Among these skills is the ability to recognize when the scale of a graph is truncated and what impression creates such a thing, as in this specific item. While some students were able to recognize that the differentiation of the two graphs was due to the different scale, they could not understand why this created a wrong impression.

> Student 41: As it is obvious at the 1st graph the crime rates ranges from 23 to 29. So, the rest of the graph is useless. On the contrary at the 2nd graph the given values are between 23 and 29 which make it more accurate.

DISCUSSION AND CONCLUSION

The goal of the present study was to assess the level of statistical literacy of pre-service teachers at their first academic year. Study findings revealed that although students had been taught Statistics in High School, their level of statistical literacy was rather low, specifically when it comes to the questioning of statistical claims in a social context. These findings are consistent with those of other studies (Godino et al., 2008) whom suggest that prospective primary school teachers in many countries enter the Departments of Education with a very limited statistical competence.

In brief, the participants were able to read and interpret simple graphs such as bar graphs, line-plots and pie-charts and had knowledge of basic statistical notions such as the mean and the range. However, they had difficulties in regards to sampling, graph evaluation and proportional thinking in a real life context. Also they could not give proper justification to statistical claims in the context of a social discussion.

For the purpose of this study we used misleading tables and graphs and we asked the participants to question claims based on them. For the answer of these specific items most of the participants relied on their experience of real world data than on their statistical knowledge. This fact indicates that even adults after the end of schooling, like students (Watson & Chick, 2004), are not able to detect the "unusual".

According to Gal (2002) there are five interrelated "knowledge bases" that must be used to exhibit statistical literacy: mathematical knowledge, statistical knowledge, knowledge of the context, literacy skills and critical questions. Analyzing the results of the present study we concluded that items with low performance demanded more than plain mathematical or statistical knowledge. We suggest that items demanding one or more of the last three "knowledge bases" (context, literacy skills and critical stand), turned out to be difficult for the pre-service teachers. Consequently, we

suggest that a content of a Mathematics Course for pre-service teachers should further focus on these aspects.

ACKNOWLEDGMENTS

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

REFERENCES

ACARA (2010). *Australian Curriculum Consultation Portal*.

Aoyama, K., Stephens, M., (2003). Graph interpretation aspects of statistical literacy: A Japanese Perspective. *Mathematics Education Research Journal*, 15(3).

Best, J., (2001). *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*. University of California Press.

Budgett, S. and Pfannkuch M. (2007) *Assessing Students' Statistical Literacy*. International Association for Statistical Education (IASE/ISI Satellite, 2007 Portugal), http://www.StatLit.org/pdf/2007BudgettPfannkuchIASE.pdf.

Cobb, P. (2000). The importance of a situated view of learning to the design of research and instruction. In J. Boaler (Ed.*), Multiple perspectives on mathematics teaching and learning* (pp. 45-82). Westport, CT: Ablex Publishing.

delMas, R., Garfield, J., Ooms, A., Chance, B., (2007). Assessing students'conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58, http://www.stat.auckland.ac.nz/serj

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1 – 25.

Ghinis, D., Korres, K., Bersimis, S. (2009). Difficulties Greek Senior High School Students Identify in Learning and the Teaching of Statistics: The Case of Experimental and Private High Schools. *Journal of Statistics Education* Volume 17, Number 3 (2009), www.amstat.org/publications/jse/v17n3/ghinis.html

Godino, J., Batanero, C., Roa, R., Wilhelmi, M., (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. . In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds), Joint ICMI/IASE Study Statistics in School Mathematics.

Kaplan, J., & Thorpe, J., (2010). Post-secondary and adult statistical literacy: Assessing beyond the classroom. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proc. of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.

Moreno, J. (2002). Toward a Statistically Literate Citizenry: What Statistics Everyone Should Know. In B. Phillips (Ed.), Proceedings of the Sixth International Conference on Teaching Statistics, CD-ROM. Cape Town, South Africa.

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

OECD. (2003). *PISA Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris, France: OECD.

Pedagogical Institute of Greece (2007). *Instructions for the syllabus of lessons in Gymnasium and Lyceum for the school year 2007–08 (Part B: Mathematics) (in Greek).* Athens: Organization of Publication of Teaching Books.

Rumsey, D. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of statistics Education,* 10(3). Online: www.amstat.org/publications/jse/v10n3/rumsey2.html

Schield, M., (2006). Statistical Literacy Survey Results: Reading Graphs and Tables of Rates and Percentages. *Conference of the International Association for Social Science Information Service and Technology (IASSIST*). www.StatLit.org/pdf/2006SchieldIASSIST.pdf

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *The second handbook of research on mathematics* (pp. 957–1010). Charlotte, NC: Information Age.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1 – 8.

Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J.Garfield (Eds.), *Handbook of Assessment in Statistics Education* (pp. 107-121).Amsterdam: IOS Press and the International Statistical Institute.

Watson, J.M. & Moritz, J.B. (2000). The development of concepts on average. Focus on Learning Problems in Mathematics, 21, 15-39.

Watson, J. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.

Watson J.M. & Chick H.L. (2004). What is unusual? The case of a media graph. *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2 (pp. 207–214).