HANDS-ON STRATEGIES IN STATISTICAL LITERACY OR EDUCATION: EFFECTS ON THE PRESENT AND FUTURE

Liliana Adriana Mendoza Saboya
PhD Economics and Business
University of Granada, Spain
lilianaadriana11@hotmail.com

The aim of this research is to identify whether use of playful ways of statistical literacy is related to their perception of using statistics in the present and future. The conceptual framework of literacy and statistical education is reviewed in a non-probabilistic sample is observed, the study population is students and teachers in Colombia. To determine the independence between the variables of the study, Chi2 p-value<0.06 is used, the results are segmented into basic and secondary, rural or urban, Educational Institutions (E. I.), by gender and age. Likewise, a portion of the sample is trained with a logistic regression model to identify the pattern of literate students who are currently using statistics and students who will not continue to study statistics after high school. This will provide clues about the early literacy strategies that can increase the use of statistics in the daily and professional life of Colombians.

STATISTICAL LITERACY AND EDUCATION

According to the introduction of Liao & Lin, (2016) the new era is now the age of knowledge based on data stored in large quantities, its management and analysis. (Gal & Murray, 2011), (Isham, 2012). As a consequence, the business world is increasing its demand for professionals in different areas who show practical skills in qualitative and quantitative information analysis. However, professionals are facing different obstacles (Cameron et al., 2017) among which stand out the inefficient literacy to face these challenges. (Zapata, 2010), (Del-Callejo-Canal et al., 2020). Therefore, it is necessary to review the statistical literacy strategies (Mclaughlin & Kang, 2017) that are being used by teachers (Hazelton et al., 2016), (Metz, 2010), undergraduate students (Bromage et al., 2021), graduate students (Kline, 2020), and even in schools for children and youth (Ben-Zvi & Makar, 2016)

The concept of literacy has been in use since the beginning of the century (Ogburn, 1940), and it has evolved (Walker, 1951), (Wallman, 1993). In the beginning of the 21st century, the most elaborate concept is the one proposed by Gal, (2002) who states that statistical literacy is "the ability to interpret, critically evaluate and communicate statistical information and messages". Taking the concepts of early literacy, which refer to children's early approach to reading and writing exercises escritura (Pérez Yglesias, 2010), (Whitehurst & Lonigan, 1998) and conventional literacy, which refers to the reading and writing exercises that are developed in school or in the formal institution. (Rugerio & Guevara, 2015), (Stahl & Yaden, 2004). In the same way, the statistical literacy has an initial literacy, which is given before the academic programs establish it for each age. Thus, the relationship between literacy and education is a complementary one. In this sense, a "literate" in statistic is a person who can develop some of the activities of daily life, aided by statistical knowledge and notation (Delmas et al., 2007).

Analogously to the definition of (UNESCO, 1958) “está alfabetizada toda persona que puede leer y escribir, comprendiéndola, una breve y sencilla exposición de hechos relativos a su vida cotidiana”, a literate statistician is defined as a person who can read and write, understanding a brief and simple statement of facts relating to his or her daily life using statistical language. In this sense, people who do not have the competencies mentioned in the basic competency standards for their age are considered illiterates who are potentially subject to statistical literacy.

In Colombia, with a length up to 1.200 hours a year, students split their time into the diverse areas of Language, Mathematics, Sciences and Citizenship Competences. As per the Colombian legislation, every Educational Institution can establish the number of hours for each area aforementioned. This organization is consigned on a document called Educational Project of the Institution. In this respect, we do not know the number of hours that elementary school, middle school, and high school students dedicate to the aleatory or probabilistic thought core (Ministry of National Education of Colombia, 1994). However, under the supposition that all the areas dedicate the same length, a quick calculation for the weekly dedication on the statistical topics would be the following: For high school, the 1.200 hours per year are divided into the 4 cores resulting in a dedication of 300

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

hours for mathematics which, being divided into the 5 types of thought we obtain 60 hours of dedication for the aleatory or probabilistic thought. This indicates that, for 40 weeks of study, students will have 1.5 hours or 90 minutes of dedication to statistical topics per week. In this time, teachers must apply more effective strategies and which they feel more identified with, which guarantee the literacy and education of their students so that they be able to read, write and understand fact of daily life by using a statistical language.

STATISTICAL LITERACY PRACTICES

Tishkovskaya & Lancaster, (2012) propose a review of statistical teaching practices, which Mafokozi, (2011) summarizes into three strategies for teaching statistics at the university: classes on the board, class experiments and group exercises within the class Although Mafokozi cannot determine whether there are significant differences in learning related to the method of teaching or literacy, his findings point to the hypothesis that the tradition (Sharma, 2015). In addition to the three methodological strategies mentioned above, we can add the playful and experimental strategies, which according to Jones & Goldring, (2017), Pérez Yglesias, (2010), Zavarrone, (2017) are a fundamental part for processes of training, exchange, and construction of knowledge in context. In this respect, the following three strategies outstand as a fundamental part of the literacy process and statistical education: Hands-on exercises (surveys, taking measurements, recording data in a format); Out-of-class experiments and Going out for practice.

As in university courses as at schools making a Hands-on exercises strategy increases the efficiency in learning. This strategy allows to share practical knowledge on mathematical concepts applied to data of the real world, which leads up to a better comprehension of concepts, and it makes students to participate more in class (Miah, 2019). Specifically, within the practical exercises for the statistics area outstand: surveys, taking measurements, recording data in a format.

On the other hand, the study made by Das, (2021) let us see that the experiments out of the classroom are a more efficient tool than the traditional ones, since this may achieve that the student shows a greater interest in knowing contents, they command with mathematical contents in places different to the classroom.

The combination between Out-of-class and in-of-class activities results efficient to improve the learning of mathematics (Song, 2020), and when the out-of-class practices are related to labs or experiments, the interest in classes could increase.

HYPOTHESIS AND METHODOLOGY

This study focuses on the evaluation of the following three playful strategies: Hands-on exercises (surveys, taking measurements, recording data in a format); Out-of-class experiments and Going out for practice, in order to identify whether the enjoyment the students have for the use of playful activities of statistical literacy is related to their perception of the use of statistics in their daily lives, in the present and to continue studying statistics after high school.

The hypothesis was evaluated in a sample of 780 students and 78 teachers between the fourth grade of elementary school to the last grade of high school (4th to 11th) from 19 educational institutions, the call to participate in this research was made in Colombia, at a national level, in public and private educational institutions. Two forms were developed with Google Forms as a mechanism to collect data, for teachers and their students.

The research was developed within the framework of the International Statistical Literacy Project - ISLP. This project works to increase the use of statistics in Colombia from an early age. The participation in the study was voluntary, we invited 5052 Educational Institutions, to whom we sent an e-mail explaining the objective of the research and the instructions for participation.

Enrollment was carried out in three stages: In the first instance, invitations were sent to the Principals of each Educational Institution, who are the highest authority within the Institution. At this stage, the principals were instructed to share the Link of the form with the teaching staff, and especially with those who were head of groups, regardless of the subject they taught. The next step was for the teachers to fill out the survey, and once it was received, the third step was to share the survey that had been designed to Students.

The survey addressed to the teachers contained 19 questions. In order to find out directly whether or not there was a difference between "statistical literacy" and "statistical education", the

following question was asked: "Do you consider that "statistical literacy" and "statistical education" are the same? Explain your answer"; the answer options were Yes or No. Likewise, the questions aimed at measuring the level of effectiveness of the teaching strategies were also included with the question: "Indicate the degree of effectiveness of the teaching strategies for statistics at your school". The response options were low, medium, and high. Finally, we included questions aimed at measuring the level of liking that teachers have for the three evaluation strategies: Hands-on exercises (surveys, taking measurements, recording data in a format); Out-of-class experiments and Going out for practice; whose response options were also low, medium and high.

The student's survey was anonymous and consisted of 20 questions. The questions were aimed at inquiring about preferences for the use of statistics, at present and in the future, and were the following, respectively: "Do you use what you know about statistics in your life outside of school? and "When you graduate from high school, would you like to continue studying statistics?", whose response options were Yes, No or Maybe. Questions aimed at measuring the level of liking for the three strategies under evaluation were also included.

In order to test the hypothesis of dependence between the similarity between "literacy" and "education"; and the playful strategies and the hypothesis of dependence between the three strategies with the use of statistics in the present and with continuing to study statistics after high school. Tests are developed with Chi2 with p-value< 0.06. and results are compared by geographic area, gender and age. Python´s Pandas package is used for the analysis (McKinney, 2010).

Although Vélez-Bedoya et al., (2021) we know that small samples are not a sufficiently optimal data set for the use of machine learning training, its use is considered necessary because it allows us to have a simulated look at the behavior of the sample. In this sense, a flow management process (Kotsiantis, 2007) is performed with the purpose of finding a supervised classification algorithm (S et al., 2020) to identify, in a multivariate way, the determinants that make a student have the perception of using statistics in the present and continuing to study statistics after high school. The algorithms evaluated were: Logistic Regression, Support Vector Machine, Decision Tree, Linear Discriminant Analysis and Quadratic Discriminant Analysis. Logistic regression was selected as the best classification algorithm option. The algorithm is evaluated through the simulation of its hyperparameters by applying the cross-validation method with 5 partitions. (Refaeilzadeh et al., 2009) (Weerts et al., 2020) The processing is performed through the GridsearchCV package. 1400 combinations with the following parameters are tested: n_components: [1 to 14]; Penalty: [l1, l2] and C = [1 to 1,00E+04]. Python´s Sklearn package is used for the analysis (Pedregosa et al., 2011). The variable used in this instrument were provided by the questions associated to gender, age (14 years-old and above), rural or Urban area, perception of the use of statistics at school, and a high level of liking of the three leisure activities under study.

For the teachers participating in this survey, "literacy" and "education" are synonymous since they both lead to the learning of knowledge. On the contrary, teachers who consider that literacy and education are not synonyms indicate that "literacy" provides basic tools to learn about a subject and, with them, to develop life skills, for example: reading and writing, while "education" is a more structured and comprehensive teaching process that goes from the general to the particular; indicating also that education is the process whose objective is to develop the intellectual capacity and understanding, for use in decision making. A non-parametric hypothesis test of independence between the categories yields the following- p-values for the three strategies: 0.1475; 0.6363 and 0.5988, which suggests that there is insufficient evidence to reject the hypothesis of independence between liking the strategies and differentiation between the concepts of literacy and education.

When comparing the perception of those who consider literacy and education to be the same or different, we find out that there is an independence between this and the effectiveness of strategies at the time of teaching statistics. Based on this independence of the teacher sample, the students' results are also inferred to be independent, given that it is these same teachers who teach the students.

There was strong evidence of a relationship between the student's geographic location with the perception that they have of using statistics in their daily life at present ($p = 0.0246$) and with continuing to study statistics in the future ($p = 0.00005$), except for the sample from the rural area and students under 14 years of age.

Table 1. Test of independence by strategy

Using statistics at present	General	Rural	Urban	Female	Male	Under 14	More 14
Continue to study statistics in the future	0,0000	0,0690	0,0001	0,0015	0,0136	0,0611	0,0004
Hands-on exercises	0,0056	0,1026	0,0574	0,0046	0,5303	0,1749	0,0056
Going out for practice	0,0003	0,3636	0,0013	0,0182	0,1731	0,2877	0,0003
Out-of-class experiments	0,0032	0,1011	0,0147	0,0766	0,0010	0,1703	0,0032

Continue to study statistics in the future	General	Rural	Urban	Female	Male	Under 14	More 14
Hands-on exercises	0,0118	0,0820	0,0464	0,0155	0,6404	0,0479	0,0119
Going out for practice	0,0136	0,3818	0,0041	0,6821	0,0005	0,0073	0,0136
Out-of-class experiments	0,0136	0,3818	0,0041	0,6821	0,0005	0,0073	0,0136

The study shows that there is dependence between the perception of the use of statistics in the present and continuing to study statistics after high school and the liking for the three proposed playful strategies. In the segmentation by rural and urban area of the student, this dependence is confirmed in this sample.

In the segmentation by gender, the female sample relates the strategy of going out to practice with the use of statistics in the present and with continuing to study statistics after high school. The strategy of practical exercises is associated with the use of statistics in the present. As for the sample of male students, there is dependence between present and future use of statistics and liking for strategy of doing experiments outside of class. Independence is evidenced between the strategy of going out to do practical work and the use of statistics in the present and with continuing to study statistics after high school.

In addition, it is evident that the strategy of having practical exercises is related to continuing to study statistics, but not to the use of statistics at present. In the case of the students´ age, the p-values are 0.4976 and 0.0008, so there is no conclusive evidence that the age of the students is related to the use of statistics in the present, but it is related to continuing to study statistics after high school. When dividing the sample at the age of 14 years, it is observed that there is a dependence relationship between age and continuing to study statistics after high school in the future (0.0001). The group of students under 14 years of age indicates independence between statistics in the present with the perception of liking for the three strategies. In the group of students older than 14 years of age, the tests of hypotheses of independence indicate that there is a relationship between liking the three strategies and the use of statistics both in the present and with continuing to study statistics after high school.

After univariate tests of independence, the result of the machine learning training yielded the following accuracy for present model: Logistic Regression: 0.64, 0.92; Support Vector: Machine: 0.64, 0.92; Decision Tree: 0.55, 0.89; Linear Discriminant Analysis: 0.64, 0.92; and Quadratic Discriminant Analysis: 0.64, 0.92; respectively. Choosing logistic regression as the best model with the following parameters C: 0.019 penalty: l2 and n_components: 6 for the variable of statistical use in the present. And with parameters C: 0.006, penalty: l1 and n_components: 3 for the variable indicating not continuing with statistics studies after high school.

After training with 80% of the sample, it is observed that the model based on the liking of literacy strategies which estimates each of the dependent variables: using statistics in the present has an accuracy of 0.61 and a precision of 0.64; and an accuracy of 0.68 and a precision of 0.33 for using statistics in the future.

CONCLUSIONS

This exercise is a first look at the evidence that literacy is a tool of education that addresses the initial and basic part of the teaching of those who begin a learning process. The teachers use strategies to teach statistics, without differentiating which of these strategies are oriented to statistical literacy or formal statistical education.

A relationship is identified between the use of statistics in the present and continuing to study statistics after school, except for the sample from the rural area, differences in this relationship by gender are not identified. The rural area does not show any relationship between the use of statistics and the different statistical literacy strategies. The female sample relates statistical literacy strategies to the current use of statistics; in a complementary way, the male sample relates the liking of these strategies with continuing to study statistics in the future.

For women, the strategy of doing practical exercises has an impact only on the present, but there is no evidence that it has an impact on the decision to continue studying statistics in the future; there is

no evidence that the strategy of going out to do experiments outside of class is related to the use of statistics in the present or to continuing to study statistics after high school.

It is evident that the use of study or literacy hands-on strategies are not sufficient to determine whether a student uses statistics in the present or will study statistics in the future. However, the use of playful strategies allows 6 out of 10 students to use statistics in their daily lives, which is a 60% literacy level. In the same way, playful strategies are determinant for the possibility that students will want to continue with statistics studies in the future.

This research presents some bases for inquiring about the factors of the decision to continue with studies in statistics after high school, and to move from statistical literacy within the initial education to formal education in statistics at the professional level.

REFERENCES

Ben-Zvi, D., & Makar, K. (2016). International Perspectives on the Teaching and Learning of Statistics. En D. Ben-Zvi & K. Makar (Eds.), *The Teaching and Learning of Statistics: International Perspectives* (pp. 1-10). Springer International Publishing. https://doi.org/10.1007/978-3-319-23470-0_1

Bromage, A., Pierce, S., Reader, T., & Compton, L. (2021). Teaching statistics to non-specialists: Challenges and strategies for success. *Journal of Further and Higher Education*, 1-16. https://doi.org/10.1080/0309877X.2021.1879744

Cameron, C., Iosua, E., Parry, M., Richards, R., & Jaye, C. (2017). More than just numbers: Challenges for professional statisticians1. *Statistics Education Research Journal*, *16*(2), 362-375. Scopus.

Celine, S., Dominic, M., & Devi, S. (2020). Logistic Regression for Employability Prediction. *International Journal of Innovative Technology and Exploring Engineering*, *9*(3), 2471-2478. https://doi.org/10.35940/ijitee.C8170.019320

Del-Callejo-Canal, D., Canal-Martínez, M., & Hákim-Krayem, M. R. (2020). Statistical thinking development in superior level students through one educative experience. *Educación Matemática*, *6*(2), 194-216. Scopus. https://doi.org/10.24844/EM3202.08

Delmas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Undefined*.

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, *70*(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., & Murray, S. T. (2011). Responding to diversity in users' statistical literacy and information needs: Institutional and educational implications. *Statistical Journal of the IAOS*, *27*(3-4), 185-195. https://doi.org/10.3233/SJI-2011-0730

Hazelton, L., Allen, M., MacLeod, T., LeBlanc, C., & Boudreau, M. (2016). Assessing clinical faculty understanding of statistical terms used to measure treatment effects and their application to teaching. *Journal of Continuing Education in the Health Professions*, *36*(4), 278-283. Scopus. https://doi.org/10.1097/CEH.0000000000000121

Isham, V. (2012). The evolving Society: United we stand. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(2), 315-335. https://doi.org/10.1111/j.1467-985X.2011.01024.x

Jones, J. S., & Goldring, J. E. (2017). Telling stories, landing planes and getting them moving—A holistic approach to developing students' statistical literacy. *Statistics Education Research Journal*, *16*(1), 102-119. Scopus.

Kline, R. B. (2020). Post p value education in graduate statistics: Preparing tomorrow's psychology researchers for a postcrisis future. *Canadian Psychology*, *61*(4), 331-341. Scopus. https://doi.org/10.1037/cap0000200

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica (Ljubljana)*, *31*.

Mafokozi, J. (2011). Nivel de alfabetización estadística del alumnado universitario de letras: El caso de la Facultad de Educación de la Universidad Complutense de Madrid. *Revista complutense de educación*, *22*(1), 95-125.

McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. 56-61. https://doi.org/10.25080/Majora-92bf1922-00a

Mclaughlin, J. E., & Kang, I. (2017). A flipped classroom model for a biostatistics short course. *Statistics Education Research Journal*, *16*(2), 441-453. Scopus.

Metz, M. L. (2010). Using GAISE and NCTM standards as frameworks for teaching probability and statistics to pre-service elementary and middle school mathematics teachers. *Journal of Statistics Education*, *18*(3), 1-27. Scopus. https://doi.org/10.1080/10691898.2010.11889585

Miah, M. (2019). *Providing hands-on experience using rapidMiner in undergraduate data mining class: Lessons learned*. *2*, 29-33. Scopus.

Ogburn, W. F. (1940). Statistical Trends. *Journal of the American Statistical Association*, *35*(209b), 252-260. https://doi.org/10.1080/01621459.1940.10500563

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825-2830.

Pérez Yglesias, M. (2010). Estrategia lúdico-creativa: Al conocimiento y la educación por el placer. *Revista Educación*, *34*(1), 55. https://doi.org/10.15517/revedu.v34i1.497

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. En L. LIU & M. T. ÖZSU (Eds.), *Encyclopedia of Database Systems* (pp. 532-538). Springer US. https://doi.org/10.1007/978-0-387-39940-9_565

Rugerio, J. P., & Guevara, Y. (2015). Alfabetización inicial y su desarrollo desde la escuela preescolar. *Ocnos: Revista de estudios sobre lectura*, *13*, 25-42. https://doi.org/10.18239/ocnos_2015.13.02

Sharma, S. (2015). Promoting Risk Taking in Mathematics Classrooms: The importance of Creating a Safe Learning Environment. *The Mathematics Enthusiast*, *12*(1), 290-306.

Stahl, S. A., & Yaden, Jr., David B. (2004). The Development of Literacy in Preschool and Primary Grades: Work by the Center for the Improvement of Early Reading Achievement. *The Elementary School Journal*, *105*(2), 141-165. https://doi.org/10.1086/428862

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical Education in the 21st Century: A Review of Challenges, Teaching Innovations and Strategies for Reform. *Journal of Statistics Education*, *20*(2), null. https://doi.org/10.1080/10691898.2012.11889641

UNESCO. (1958). *Conferencia General*. https://unesdoc.unesco.org/ark:/48223/pf0000114584_spa/PDF/114584spao.pdf.multi

Vélez-Bedoya, Á. R., Mendoza-Saboyá, L. A., & Luna-Eraso, J. L. (2021). Determination of Competitive Management Perception in Family Business Leaders Using Data Mining. En *New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques* (pp. 435-462). Springer International Publishing. https://doi.org/10.1007/978-3-030-71115-3_19

Walker, H. M. (1951). Statistical Literacy in the Social Sciences. *The American Statistician*, *5*(1), 6-12. https://doi.org/10.2307/2685917

Wallman, K. K. (1993). Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association*, *88*(421), 1-8. https://doi.org/10.2307/2290686

Weerts, H., Mueller, A., & Vanschoren, J. (2020). *Importance of Tuning Hyperparameters of Machine Learning Algorithms*.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development*, *69*(3), 848-872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

Zapata, L. (2010). ¿Cómo contribuir a la alfabetización estadística? *Memoria 11a: Encuentro Colombiano de Matemáticas*, 9.

Zavarrone, E. (2017). Latent Growth and Statistical Literacy. En N. C. Lauro, E. Amaturo, M. G. Grassia, B. Aragona, & M. Marino (Eds.), *Data Science and Social Research* (pp. 291-300). Springer International Publishing. https://doi.org/10.1007/978-3-319-55477-8_27