Design of statistical projects at the school level, advances and challenges when engaging in the statistical problem-solving process

Francisco Javier Trejo Moreno and Eleazar Silvestre Castro
University of Sonora, Mexico
a220208612@unison.mx

This paper presents the results of a design-based research study aimed at fostering 8th-grade students' Statistical Thinking through contextualized investigations. Drawing on GAISE II recommendations for novice levels (A and B) and the Statistical Working Spaces framework, two statistical projects were designed using local orange production as context. Findings indicate that students made significant progress in anticipating variability, organizing data, and applying measures of central tendency and variability. Nevertheless, difficulties were observed when formulating comparative questions, understanding concepts such as sample and variable, and interpreting graphical representations like dot plots and histograms.

INTRODUCTION

Since the late 20th century, statistical projects and the use of real data have gained increasing relevance in international recommendations for statistics education (Cobb, 1992; Garfield & Gal, 1999; Batanero, 2000). Their primary purpose is to provide students with authentic opportunities to engage in statistical inquiry (Davidson, 2023), which requires activating concepts and processes that go beyond memorizing definitions or procedures (Silva & Pinto, 2014). In doing so, students connect statistical ideas with contextualized situations, helping to bridge the gap between theory and practice (Ghinis et al., 2005).

In this regard, statistical projects have been recognized as an effective strategy to foster investigative processes in the classroom and to support the development of Statistical Thinking (ST) (Aliaga et al., 2005; Carver et al., 2016). Nevertheless, Gómez-Blancarte and Santana (2018) conducted a literature review of 67 studies based on sources such as international research journals, conference proceedings, graduate theses, and online databases. Their findings report that around 80% of these proposals employ projects primarily to showcase previously acquired knowledge rather than to explore new curricular content (Santana, 2020). This prevailing approach limits their potential as vehicles for inquiry, revealing a gap between their pedagogical potential and their actual use in classroom practice.

This gap is also evident in the Mexican curriculum. Although recent reforms emphasize project-based learning within a STEAM context to promote authentic and interdisciplinary experiences in Mathematics and Statistics (Secretaría de Educación Pública, 2024), curricular materials frequently present statistical concepts in isolated and decontextualized ways. Tasks often prioritize computation over inquiry or interpretation, thereby limiting students' opportunities to engage in meaningful statistical investigations (Silvestre & Trejo, 2024).

To address this issue, this study explores the implementation of a design-based research project with 8th-grade students, focused on fostering ST through statistical projects. The research examines how students navigate different phases of the investigative cycle, highlighting both the challenges they face and the progress they make. Therefore, the study is guided by the following research question: How do statistical projects in secondary school support the development of students' ST, and what challenges emerge throughout the process?

FRAMEWORK

In this study, Statistical Thinking (ST) is understood as the mobilization of concepts and procedures throughout the phases of a statistical investigation, aimed at making decisions that consider the variability of the phenomenon under study (Wild & Pfannkuch, 1999; delMas, 2004; Gómez-Blancarte et al., 2021).

To support the development of ST, the design of statistical projects was guided by the objectives outlined in the Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) framework (Bargagliotti et al., 2020). GAISE II emphasizes solving statistical problems through a structured investigative cycle composed of four phases. Each phase includes objectives intended to

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

support students' progression in developing ST, providing a clear pathway for teaching and learning statistics. The projects targeted levels A and B of the framework, where students are introduced to exploratory data analysis and descriptive statistics, fostering an understanding of variability and contextual interpretation. Level C was excluded as it addresses inferential concepts beyond the scope of the secondary school curriculum.

Additionally, the Statistical Working Spaces (SWS) model was incorporated to structure students' actions and their use of statistical objects throughout the projects (see Figure 1). This model conceptualizes students' statistical activity through the interaction of three forms of genesis: semiotic genesis, involving the construction and interpretation of representations; instrumental genesis, concerning the use of tools and procedures; and discursive genesis, which connects statistical concepts with their contextual or real-world meaning (Kuzniak et al., 2016; Vidal-Szabó et al., 2020).

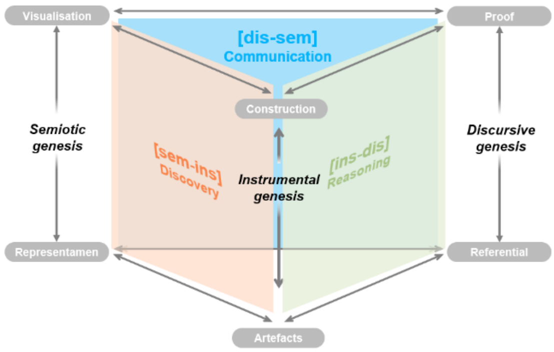

Figure 1. Statistical Working Space, adapted from Vidal-Szabó et al., (2020).

METHODOLOGY

This study followed a design-based research methodology (Molina, 2021), involving iterative cycles of instructional design, implementation, and analysis aimed at fostering ST in middle school students through contextualized projects.

Participants and Context

The research was conducted by the first author with a group of 24 eighth-grade students (aged 13–14) from a public secondary school in Hermosillo, Mexico. The statistics curriculum at this level includes two key contents: "Data Collection and Representation" and "Interpretation of Information through Measures of Central Tendency and Spread" (Secretaría de Educación Pública, 2024), which framed the design of the projects.

Resources

In alignment with the Mexican curriculum, the projects integrated elements from students' immediate environment to promote teaching and learning through authentic, community-relevant situations. Given that Hermosillo is a citrus-producing region, the study of oranges from Costa de Hermosillo was selected as the context for the design of two statistical projects, each consisting of five activities. Figure 2 summarizes how these activities are distributed according to GAISE phases.

The first project investigated variability in the number of seeds between two orange varieties (Regional Temprana and Valencia Tardía) to determine their suitability for fresh consumption. Students anticipated variability, collected seed count data, and organized it using graphical representations. They analyzed the seed counts by calculating the mean, mode, range, minimum, and maximum values. Finally, students compared these statistics alongside dot plots to recommend the most appropriate variety for consumption.

The second project examined variability in juice percentage for the same two orange varieties to determine which was better suited for juice production. Students anticipated variability in orange weight and juice content, formulated a comparative statistical question, and collected data. As with seed counts, they calculated the juice percentage for each orange and discussed methods to summarize and represent this variable using frequency tables and histograms. Data analysis involved calculating the mean, median, mode, range, minimum, and maximum values for juice percentage. In the final phase, students compared these statistics and graphical representations to identify which variety yielded more juice, thus addressing the project's statistical question.

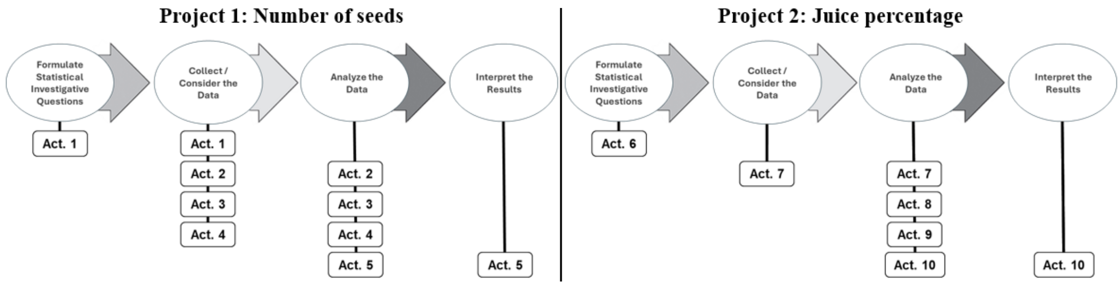

Figure 2. Progression of activities of the first and second project.

Data Collection and Analysis

Data collection took place during a 25-day implementation period and included students' worksheets, classroom observations, and audio recordings of group discussions. All responses generated during activities were analyzed to provide a general evaluation of the project. The number of responses per activity varied according to student attendance in each session. The analysis focused on identifying evidence of students' progression in their ST and the challenges they faced while solving the activities.

RESULTS

Results were organized according to the GAISE phases. In each phase, students' responses were examined based on the activity objectives. Additionally, the SWS model was employed to interpret students' statistical activity, focusing on their actions, representations, use of tools, and the arguments provided within the context of the projects. Below, we present some student responses that illustrate aspects of our findings.

Phase I: Formulate Statistical Investigative Questions

In the first project's initial activities, 10 out of 19 students successfully anticipated variability in the number of seeds by proposing intervals rather than single values. For example:

Student 9: 5 to 8 [seeds].

Student 11: from 3 to 6 [seeds].

By the time they reached the second project, 20 out of 20 students were able to anticipate variability, predicting differences in weight and juice content. This was reflected in the students' responses that recognized how multiple characteristics could influence the outcome.

Student 1: *No, just because they have the same weight doesn't mean they have the same juice, because they can also have a different number of seeds or a thicker shell.*

This progression aligns with GAISE II recommendations, which emphasize the importance of anticipating variability as a necessary step before formulating statistical questions. However, notable difficulties emerged when students attempted to construct comparative questions involving samples and populations. Only 8 out of 20 students were able to formulate a question; while one exhibited a comparative structure, none explicitly referred to the variable of interest (juice percentage) or populations involved (Regional Temprana and Valencia Tardía):

Student 2: How much juice should an orange have?

Student 6: Is the taste of orange the same?

When asked how many oranges would be needed to identify the best variety for juice production, most students appeared uncertain and could not propose a sample size. Only 4 out of 20 participated in this discussion, and their responses varied widely. Two suggested very small samples, while the other two proposed more reasonable values, such as 50 or 100, recognizing that reliable comparisons require larger sample sizes:

Student 22: 2 [oranges] (meaning 1 orange per variety).

Student 5: A small amount because they are very different, 5 or 6 [oranges].

Overall, students demonstrated clear progress in anticipating variability; however, formulating comparative statistical questions and determining appropriate sample sizes remained challenging.

Phase II: Collect / Consider the data

In this phase of the first project, students were provided with a sack of oranges and instructed to record the number of seeds in each fruit. They organized their data in various formats and discussed which methods of organization were most suitable, recognizing that data can be structured in multiple ways depending on the purpose. These activities activated both semiotic and instrumental genesis processes, as students designed data collection tools, coded information, and interpreted diverse representations. Figure 3 shows examples of instruments created by four students.

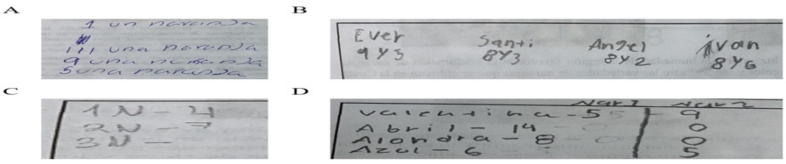

Figure 3. Instruments designed by 5 students.

A notable pattern during this phase was the students' tendency to retain the name of the person who collected each data point, a practice identified in 16 of the 22 instruments produced. We refer to this as *tracking the data*, a behavior that reflects an early stage of ST in which learners emphasize the provenance of individual observations over their abstraction.

While students showed procedural progress in data collection, conceptual difficulties emerged when reasoning about the limitations and nature of the data. For instance, although all students acknowledged that analyzing only ten oranges was insufficient to represent the population, most (14 out of 16) attributed this limitation solely to variability among oranges, overlooking the implications of using a small sample size. Only two students explicitly referred to the need for a larger sample to adequately characterize orange production in the region. Two examples illustrating this are:

Student 7: No, because the number of seeds varies.

Student 20: No, because it would have to be a larger number [of oranges] to analyze.

Phase III: Analyze the data

In the first project, students constructed pictograms, frequency tables, tally tables and dot plots to represent variability in the number of seeds (see Figure 4). In the second project, they worked with grouped frequency tables and histograms to visualize the distribution of juice percentages. Additionally, they described variability in data distributions by identifying measures of central tendency (mean, median, and mode), a variability measure (range) and the shape of the distribution (clusters, gaps, and symmetry). These activities fostered the construction, coding, and interpretation of information through multiple representations, activating both semiotic and instrumental genesis processes.

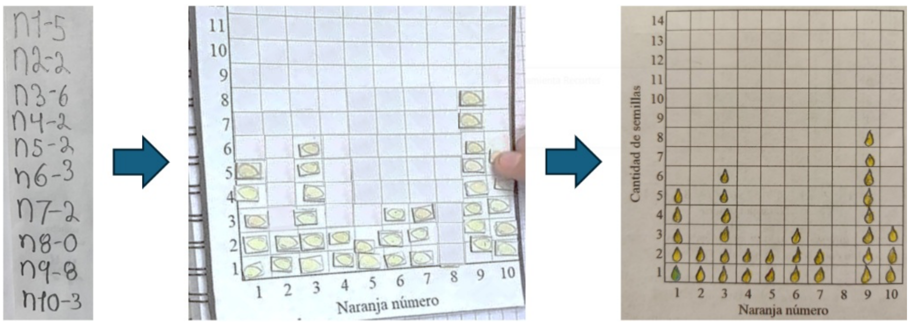

Figure 4. Seed distribution in a sack of oranges.

While students showed notable progress in representing data and calculating descriptive statistics, difficulties emerged when they were required to use these representations to compare or contrast two distributions based on their graphical and numerical summaries (see Figure 5). These tasks demanded applying previously learned procedures and activating discursive genesis, a process that proved particularly challenging for students. Specifically, only 10 out of 17 students successfully used the calculated statistics to support their arguments. Among these, 7 students used the range as the main comparative criterion, while 3 focused on the count of seedless oranges shown in dot plots.

> Student 16: Regional Temprana, because it goes from 0 to 7 [seeds], and in Valencia Tardia, it goes from 0 to 14 [seeds].
>
> Student 19: Regional Temprana, because it has more [oranges with] zero seeds.

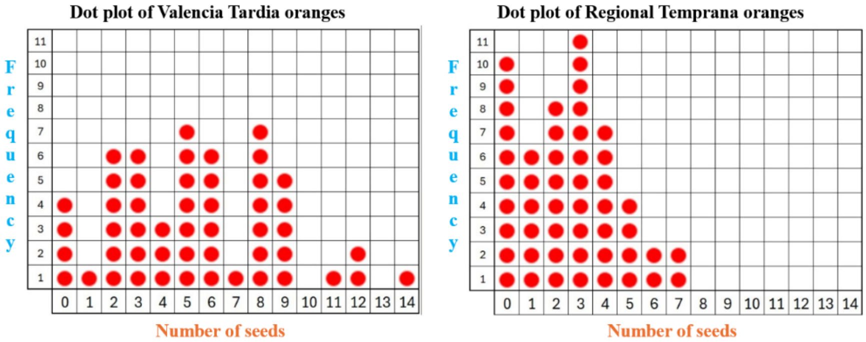

Figure 5. Dot plots of both orange varieties.

Of the remaining students, four mentioned that, during data collection, one orange variety had more seeds than the other without referring to statistical summaries. Another three stated that both varieties had the same number of seeds, possibly because they counted the number of points in the dot plot rather than the actual seed counts. These results confirm that, while procedural competencies improved, the ability to interpret and use statistical information to justify conclusions remained a demanding aspect of students' ST.

> Student 17: I think they have the same number [of seeds].
>
> Student 21: Temprana, because when we analyzed them, Regional Temprana had fewer seeds than Valencia Tardía.

Phase IV: Interpret the results

In the final phase of the first project, students employed the minimum, maximum, range, mean, and mode to answer the investigative question regarding the suitability of each orange variety for fresh consumption. Although 15 out of 18 students correctly used these statistics, none incorporated information from the dot plots when presenting their conclusions. This suggests that, while discursive genesis was activated using numerical summaries as arguments, students struggled to integrate graphical representations as complementary evidence.

> Student 18: The range for Temprana Regional is 7 and goes from 0 to 7, and the most likely number of seeds is 3
>
> Student 9: Regional Temprana, the minimum and maximum number of seeds is 0 and 7. The range is 7. The most frequent number of seeds is 3. The average number of seeds is 2.58.

This limitation became more evident in the final phase of the second project, particularly when students worked without teacher guidance. In this case, they encountered difficulties interpreting histograms: all 19 students misidentified the variety with the highest juice percentage, and none determined the modal or median intervals as requested. Although two students attempted to compare ranges between orange varieties, both reached incorrect conclusions. These difficulties point to weaknesses in semiotic genesis, as students struggled to interpret and assign meaning to the histograms they had produced (see Figure 6).

> Student 18: Regional Temprana, because it has a higher juice percentage. Regional Temprana (32.813 to 62.921) and Valencia Tardía (42.254 to 62.857) […] With the same number of oranges, more juice was obtained from Regional Temprana.
>
> Student 9: Regional Temprana, because even though only one percent [one orange] has 62.921%, it could be the one you get, and that's very good.
>
> Student 15: I think Regional Temprana because it has more [oranges] on the left side.

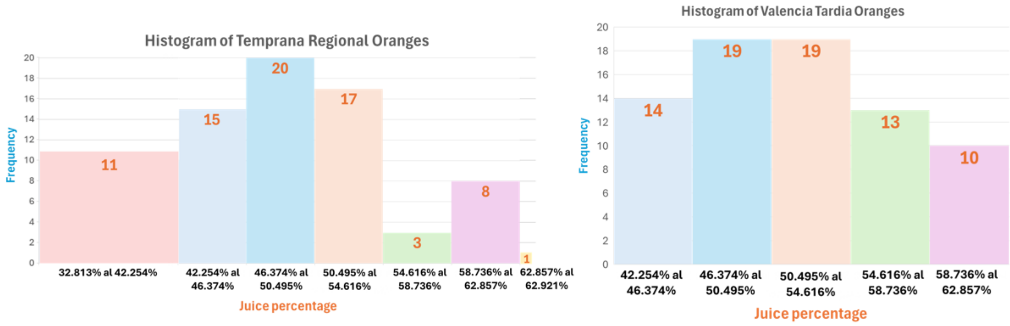

Figure 6. Histograms of both orange varieties.

CONCLUSION AND DISCUSSION

The results of this study offer insights into both the potential and the challenges of engaging middle school students in the statistical problem-solving process. Students demonstrated progress in key areas: they developed an awareness of data variability, improved their ability to organize and represent data, and applied measures of central tendency and variability to compare two orange varieties. Nonetheless, persistent difficulties emerged in aspects requiring more abstract and conceptual understanding. These included formulating comparative statistical questions involving samples and populations, estimating appropriate sample sizes, and integrating multiple forms of evidence when constructing arguments. Addressing these challenges will require instructional sequences that place greater emphasis on these topics, guiding future iterations of this design.

During the first phase, students struggled to formulate comparative statistical questions involving samples and populations. Research on this challenge remains limited; Arnold (2018) highlights the lack of consensus on what defines a “good” statistical question, while other studies report difficulties students face in linking data analysis to the investigative question (Pfannkuch & Horring, 2005) and in generating questions that guide inquiry (Burgess, 2007). Students generally have little to no prior experience in formulating statistical questions, which further complicates their performance in this phase. In the data collection phase, students recognized that their sample was not representative of the population; however, their reasoning focused exclusively on variability among individual cases, without translating this recognition into proposals for an adequate sample size. This reflects a limited understanding of sample representativeness and the role of sample size, issues well documented in the literature (e.g., Makar & Rubin, 2009; Silvestre et al., 2022) but rarely examined in relation to the earliest stages of the statistical problem-solving process.

In the data analysis and interpretation phases, students showed increasing proficiency in constructing data representations and calculating measures of central tendency and variability. Yet, they often treated numerical and graphical outputs as independent rather than complementary, limiting their capacity to draw conclusions from the full range of evidence. Moreover, when interpreting dot plots and histograms, many adopted a local view, focusing on individual data points, rather than perceiving the distribution as an aggregate, a difficulty already reported in the literature (e.g., Hancock et al., 1992; Ben-Zvi & Arcavi, 2001; Makar & Confrey, 2005; Biehler et al., 2018). Strengthening activities within the reasoning plane of the SWS model may help students coordinate multiple representations more effectively when interpreting results in the final phase of the GAISE framework.

In terms of educational implications, this study reaffirms the pedagogical value of project-based approaches for fostering students' ST. By engaging students in the analysis of both discrete (number of seeds) and continuous variables (percentage of juice) within the same thematic focus, these projects demonstrated the relevance and potential of project-based learning in analyzing diverse types of data. Supporting teachers with resources and professional development opportunities remains essential to facilitate the design and implementation of interdisciplinary projects that promote students’ ST (Flores & Pinto, 2017; Espino-Flores et al., 2023).

REFERENCES

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P. & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College Report.* American Statistical Association.

Arnold, P., and Pfannkuch, M. (2018). Critiquing Investigative Questions. In A. Sorto, A. White & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*, ISI/IASE.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L. & Spangler, D. (2020). *The Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II).* American Statistical Association and National Council of Teachers of Mathematics.

Batanero, C. (2000). ¿Hacia dónde va la educación estadística? *Blaix, 15*, 2–13. https://www.ugr.es/~batanero/pages/ARTICULOS/BLAIX.pdf

Ben-Zvi, D. & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics 45*, 35–65 https://doi.org/10.1023/A:1013809201228

Biehler, R., Frischemeier, D., Reading, C. & Shaughnessy, J. (2018). Reasoning about data. In D. Ben-Zvi, K. Makar & J. Garfield (Eds), *International Handbook of Research in Statistics Education* (pp. 139-192). Springer. https://doi.org/10.1007/978-3-319-66195-7_5

Burgess, T. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* [Doctoral thesis, Massey University].

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G., Velleman, P., Witmer, J. & Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report.* American Statistical Association.

Cobb, G. (1992). Teaching Statistics. In L. Steen (Eds.), *Heeding the Call for Change: Suggestions for Curricular Action* (pp. 3 - 43), The Mathematical Association of America.

Davidson, A. (2023). A Review of the Use of Investigative Projects in Statistics and Data Science Courses. *Journal of Statistics and Data Science Education, 32*(2), 188–201. https://doi.org/10.1080/26939169.2023.2240385

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J.Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79–95). Kluwer Academic Publishers.

Espino-Flores, G., Gómez-Blancarte, A. & Inzunza, S. (2023). Exploración del desarrollo de competencia interdisciplinar en formadores de docentes mediante el diseño de proyectos estadísticos. *Inter Disciplina*, *11*(29), 333–357. https://doi.org/10.22201/ceiich.24485705e.2023.29.84494

Flores, A., & Pinto, J. (2017). Características de la enseñanza de la estadística por proyectos. *Acta Latinoamericana de Matemática Educativa*. 263–271.

Garfield, J., y Gal, I. (1999). Assessment and Statistics Education: Current Challenges and Dircections. *International Statistical Review, 67*(1), 1–12. https://doi.org/10.1111/j.1751-5823.1999.tb00377.x

Ghinis, D., Chadjipantelis, T. & Bersimis, S. (2005). Experiences from Teaching Statistics Using Directed Projects in Greek Elementary School. *Teaching Statistics, 27*(1), 2–7. https://doi.org/10.1111/j.1467-9639.2005.00188.x

Gómez-Blancarte, A. & Santana, A. (2018). Research on statistical projects: looking for the development of statistical literacy, reasoning and thinking. In A. Sorto, A. White & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*, ISI/IASE.

Gómez-Blancarte, A., Chávez, R. & Aguilar, R. (2021). Un cuestionario sobre la enseñanza de cultura, razonamiento y pensamiento estadístico: la enseñanza de los profesores de educación media superior en México. *Statistics Education Research Journal, 20*(2). https://doi.org/10.52041/SERJ.V20I2.397

Hancock, C., Kaput, J. & Goldsmith, L. (1992). Authentic Enquiry with Data: Critical Barriers to Classroom Implementation. *Educational Psychologist, 27*(3), 337–364.

Kuzniak, A., Tanguay, D. & Elia, I. (2016). Mathematical Working Spaces in schooling: an introduction. *ZDM Mathematics Education, 48*(6), 721–737. https://doi.org/10.1007/s11858-016-0812-x

Makar, K. & Confrey, J. (2005). Variation-talk: Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54. https://doi.org/10.52041/serj.v4i1.524

Makar, K. & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105. https://doi.org/10.52041/serj.v8i1.457

Molina, M. (2021). Investigación de diseño educativa: un marco metodológico en evolución. En Diago, P. D., Yáñez D. F., González-Astudillo, M. T. y Carrillo, D. (Eds.), *Investigación en Educación Matemática XXIV* (pp. 83–97). SEIEM.

Pfannkuch, M. & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Round Table* (pp. 204–218). International Statistical Institute.

Santana, A. (2020). Un estudio con futuros profesores de telesecundaria sobre el aprendizaje de estadística basado en proyectos [Doctoral thesis]. https://doi.org/10.13140/RG.2.2.23230.61766

Secretaría de Educación Pública. (2024). *Programa de estudio para la educación secundaria: Programa Sintético de la Fase 6*.

Silva, M. & Pinto, S. (2014). Teaching statistics through learning projects. *Statistics Education Research Journal*, *13*(2), 177–186. https://doi.org/10.52041/serj.v13i2.289

Silvestre, E. & Trejo, F. (2024). Exploración del currículum estadístico del nivel secundaria: el caso de la Nueva Escuela Mexicana. In G. Ramírez & N. Obregón. (Eds.), *Memorias VIII Encuentro sobre Didáctica de la Estadística, la Probabilidad y el Análisis de Datos Cartago, Costa Rica.*

Silvestre, E., Urrea, M. & Medina, M. (2022). How do tertiary students assess sampling on social contexts and media reports? Exploring students' statistical literacy with relation to sampling. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics.* ISI/IASE

Vidal-Szabó, P., Kuzniak, A., Estrella, S. & Montoya, E. (2020). Análisis cualitativo de un aprendizaje estadístico temprano con la mirada de los espacios de trabajo matemático orientado por el ciclo investigativo. *Revista Educación Matemática*, *32*(2), 217-246. https://doi.org/10.24844/EM3202.09

Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.2307/1403699