http://doi.org/10.14393/ER-v30a2023-32

Double-entry tables: reading and calculating probabilities by secondary school students[1]

Daniela Calderón Torres[2]
Jaime I. García-García[3]
Nicolás Fernández Coronado[4]
Elizabeth H. Arredondo[5]

ABSTRACT

The double-entry table is a useful representation for the analysis of two statistical variables, whose teaching is essential in the student's education. This qualitative research, descriptive-exploratory, aimed to analyze the reading and calculation of probabilities in double-entry tables carried out by high school students. Using the content analysis technique, the responses of 75 students to two reading and three probability calculation tasks are analyzed. The results show that the majority of the students have mastery of the first levels of Curcio's reading and make use of Laplace's rule, or the rule of three, as a solution strategy; however, some have errors, for example, they confuse a conditional probability with its inverse. The above evidence the need to design teaching proposals that address the various mathematical objects and processes related to the reading and calculation of probabilities in double-entry tables.

KEYWORDS: Strategy. Error. Reading levels. Probability calculation. Contingency tables.

---

[1] English version by Nicolás Alonso Fernández Coronado. *E-mail:* nicolasalonso.fernandez@alumnos.ulagos.cl.

[2] Professor of Secondary Education, mention in Mathematics and Computer Science. Universidad de Los Lagos, Osorno, Chile. Orcid: https://orcid.org/0000-0001-7036-3454. *E-mail*: danieladelpilar.calderon2@alumnos.ulagos.cl.
[3] D. in Science, specializing in Educational Mathematics. Universidad Metropolitana de Ciencias de la Educación, Santiago, Chile. Orcid: https://orcid.org/0000-0002-8799-5981. *E-mail*: jaime.garcia@umce.cl.
[4] Student of Master's program in Mathematics Education. Universidad de Los Lagos, Osorno, Chile. Orcid: https://orcid.org/0000-0002-9613-3144. *E-mail*: nicolasalonso.fernandez@alumnos.ulagos.cl.
[5] D. in Science, specializing in Educational Mathematics. Universidad de Los Lagos, Osorno, Chile. Orcid: https://orcid.org/0000-0002-5285-1603. *E-mail*: elizabeth.hernandez@ulagos.cl.

http://doi.org/10.14393/ER-v30a2023-32

Tabelas de dupla entrada: leitura e cálculo de probabilidades por alunos do ensino médio

RESUMO
A tabela de dupla entrada é uma representação útil para a análise de duas variáveis estatísticas, cujo ensino é essencial na formação do aluno. Esta investigação qualitativa, de carácter descritivo-exploratório, teve como objetivo analisar a leitura e o cálculo de probabilidades em tabelas de dupla entrada efetuados por alunos do ensino médio. Utilizando a técnica de análise de conteúdo, foram analisadas as respostas de 75 alunos a duas tarefas de leitura e a três tarefas de cálculo de probabilidades. Os resultados mostram que a maioria dos alunos domina os primeiros níveis de leitura de Curcio e utiliza a regra de Laplace, ou a regra de três, como estratégia de solução; no entanto, alguns cometem erros, por exemplo, confundem uma probabilidade condicional com a sua inversa. Isto mostra a necessidade de conceber propostas de ensino que abordem os diferentes objetos e processos matemáticos em torno da leitura e cálculo de probabilidades em tabelas de dupla entrada.
PALAVRAS-CHAVE: Estratégia. Erro. Nível de leitura. Cálculo de probabilidades. Tabelas de contingência.

Tablas de doble entrada: lectura y cálculo de probabilidades por estudiantes de educación media

RESUMEN
La tabla de doble entrada es una representación útil para el análisis de dos variables estadísticas, cuya enseñanza es esencial en la formación del estudiante. Esta investigación cualitativa, descriptiva-exploratoria, tuvo por objetivo analizar la lectura y el cálculo de probabilidades en tablas de doble entrada realizado por estudiantes de educación media. Bajo la técnica de análisis de contenido, se analizan las respuestas de 75 estudiantes a dos tareas de lectura y tres de cálculo de probabilidades. Los resultados muestran que la mayoría del estudiantado posee dominio de los primeros niveles de lectura de Curcio y hace uso de la regla de Laplace, o la regla de tres, como estrategia de solución; sin embargo, algunos presentan errores, por ejemplo, confunden una probabilidad

condicional con su inversa. Lo anterior evidencia la necesidad de diseñar propuestas de enseñanza donde se aborden los diversos objetos y procesos matemáticos en torno a la lectura y cálculo de probabilidades en tablas de doble entrada.

PALABRAS CLAVE: Estrategia. Error. Nivel de lectura. Cálculo de probabilidades. Tablas de contingencia.

* * *

Introduction and antecedents

In recent years, the teaching of probability and statistics has become relevant because it provides students with knowledge and skills to interpret probabilistic messages, understand statistical information represented in tables or graphs, check conjectures and make decisions based on data analysis; in other words, it allows them to face and find solutions to everyday situations where randomness and variability are present.

In the didactics of probability and statistics, for the last two decades there has been a growing body of research focused on the terms known as statistical literacy and probabilistic literacy (GAL, 2002; 2005). These terms consist of: 1) the ability to interpret and critically evaluate statistical or probabilistic information from different contexts and, thus, make informed decisions; and 2) the ability to formulate, discuss or communicate opinions regarding such information. Several authors (GAL, 2019; RODRÍGUEZ-ALVEAL et al., 2018) point out both literacies as relevant needs in the education of students and in teacher formation, with the purpose of forming statistically and probabilistically literate citizens.

Gal (2005) points out the calculation of probabilities as an essential element of probabilistic literacy, referring to the ways to calculate or estimate the probability of events; being the classic, frequential and subjective approaches to probability, those that are promoted in the school mathematics curriculum of several countries (such as Chile and Mexico) for the calculation or estimation of probabilities (GAL, 2005; SÁNCHEZ, 2009).

With respect to the statistical table, it is useful for the organization, description and analysis of data. Schield (2006) points out that a statistically literate person should critically read the statistical tables that are commonly found in their work environment, in the press, etc.; but not only read them in a literal way, but also identify the variability and trend in the data, and the possible errors that may distort the information that is intended to be communicated.

The double-entry table is a representation used to record and present the frequency distribution of a two-dimensional statistical variable. Its usefulness for the study of qualitative data, its presence in the media and the world of work, as well as its use in tasks related to the calculation of probabilities, make its reading and interpretation basic elements of statistical and probabilistic literacy (ESTRADA; DÍAZ, 2007); under this perspective, its teaching is considered essential in the student's education. However, several authors identify basic levels with respect to the understanding of statistical tables and errors in the calculation of probabilities in double-entry tables. Some studies related to this topic are mentioned below.

In relation to the reading of statistical tables, Rodríguez and Sandoval (2012) analyze the reading and construction of graphs and tables achieved by 47 Chilean basic education teachers in service and 44 in training, identifying that both groups present basic reading skills, meaning that they are located in Curcio's reading level 1, reading the data, with only some participants reaching level 2, reading within the data. Díaz-Levicoy et al. (2016) analyze the reading level of statistical tables reached by 121 Chilean female teachers in training in early childhood education, finding that most of the participants master levels 1 and 2, reading the data and reading within the data; however, only one third of the participants reach level 3, reading beyond the data, by

http://doi.org/10.14393/ER-v30a2023-32

predicting some data or trend from the information delivered in the table. For their part, García-García et al. (2019) conducted a comparative analysis on the level of comprehension of a statistical table between Mexican and Chilean university students (36 and 35, respectively), considering a hierarchy proposed from the articulation of Curcio's levels and Aoyama's levels; their results show that most students from both countries reach level 2, comparative; however, Mexican students present greater errors in the comparison of data. Gea et al. (2020) analyze the responses of 69 future elementary school teachers to a task involving the construction and reading of a double-entry table; with respect to reading, participants show difficulties in identifying joint and conditional frequencies involving probability calculations.

Regarding the calculation of probabilities in double-entry tables, Díaz and De la Fuente (2005) analyze the difficulties presented by 154 psychologists in training, identifying that, although most of them correctly calculate simple probability, few calculate conditional and compound probability, identifying that participants confuse an event with its complement and probability with favorable cases. Estrada and Díaz (2006; 2007) analyze the semiotic conflicts presented by future teachers when calculating probabilities in double-entry tables, with results similar to Díaz and De la Fuente (2005) regarding the handling of probabilities, and identifying semiotic conflicts such as confusing a conditional probability with its inverse, confusing a conditional probability with a joint probability, and confusing an event with its complement. Contreras et al. (2010) analyzed the responses of 69 teachers in training in primary education with respect to the calculation of probabilities (simple, compound and conditional) in double-entry tables. In comparison with the results of Estrada and Díaz (2006), there is a low percentage of correct answers and a higher proportion of participants who do not address the requested tasks, while, among the semiotic conflicts reported, the confusion of the probabilities with its inverse stands out.

The results of these studies, carried out with university students and teachers, evidence low reading levels and present errors in the calculation of probabilities, which correspond to challenges in the formation of statistically and probabilistically literate students. From this perspective, with the purpose of exploring this line of research at another educational level and to present an overview that facilitates addressing errors at lower educational levels, our objective is to analyze the reading and calculation of probabilities in double-entry tables carried out by Chilean high school students.

Semiotic analysis of the double entry table

The double-entry table is a table, structured by columns and rows, whose purpose is to summarize information about two statistical variables and to record the joint distribution of these variables (GEA et al., 2020). It contains as many rows and columns as there are categories of the two statistical variables that constitute it. Because several concepts and their interrelationships underlie this representation, the double-entry table is considered a complex semiotic object (GEA et al., 2020), whose simplest form is when the variables have only two categories (ESTRADA; DÍAZ, 2007) (see Table 1).

TABLE 1: Format of the simplest form of a double-entry table

	A	**no A**	**Total**
B	a	b	a + b
no B	c	d	c + d
Total	a + c	b + d	a + b + c + d

Source: Adapted from Estrada and Diaz (2007)

From this table, it is possible to derive the following types of frequencies:

- **Marginal absolute frequencies:** quantities in the right column and in the bottom row, therefore, they can be by rows (a+b and c+d) and by columns (a+c and b+d).
- **Double absolute frequencies:** quantities of the four central cells (a, b, c and d) and indicate the frequency at which specific intersections of values of the variables occur.
- **Conditional absolute frequencies:** Correspond to the frequency for the value of one variable, leaving fixed a value of the other. Mathematically, these conditional absolute frequencies are equal to the double ones; however, they are not perceived psychologically in the same way, since their reading attends to the indicated condition, that is, in relation to different values (the total quantity of which it is part is not the same) (GEA et al., 2020).

From these absolute frequencies, the marginal, double and conditional relative frequencies can be determined (double relative frequencies do not coincide with conditional relative frequencies); and assuming equiprobability of all cases in the sample, the associated probabilities can be calculated (ESTRADA; DÍAZ, 2007; GEA et al., 2020).

Curcio's Taxonomy

To analyze the reading of double-entry tables carried out by middle school students, Curcio's taxonomy was considered (CURCIO, 1989; FRIEL; CURCIO; BRIGHT, 2001). This taxonomy was established for the reading of statistical graphs; however, it has been adapted and used for the analysis of the reading of statistical tables (DÍAZ-LEVICOY et al., 2016; GARCÍA-GARCÍA et al., 2019):

http://doi.org/10.14393/ER-v30a2023-32

- Level 1. Read data: corresponds to the literal reading of data from the double-entry table, without interpretation or additional calculations.
- Level 2. Reading within the data: corresponds to the interpretation and integration of data from the double entry table, which involves comparing data or applying simple mathematical calculations.
- Level 3. Reading beyond the data: corresponds to making some inference or prediction from the data about information not explicit in the double-entry table.
- Level 4. Reading behind the data: corresponds to the critical assessment of the information represented in the double-entry table (validity and reliability), the collection and organization of the data, the interpretations or conclusions made by another person, among other considerations.

This research focuses on the analysis of the proficiency of reading levels 1 and 2 of middle school students; these levels are related, respectively, to the reading of double and marginal absolute frequency (when not shown in the double-entry table). This is because we consider both levels as essential for the calculation of probabilities in this type of statistical representation.

Error and strategy

The study of errors is a topic of interest in mathematics education research, because it helps to explain part of the problem of learning mathematics (RICO, 1996; SOCAS, 1997). According to Socas (1997), an error is generated by difficulties and obstacles in the pedagogical process and an inadequate cognitive scheme in the student, and is understood as the external manifestation (written, verbal, among others) of these, for example, in the use of mathematical symbols or algorithms.

On the other hand, according to Poggioli (1999), a problem-solving strategy is understood as the mental operation used by the student to think about the representation of goals and data, with the purpose of transforming them and obtaining a solution; this strategy includes heuristic methods, algorithms and divergent thinking processes.

Under this perspective, in this research we focus on the analysis of the errors and strategies manifested by high school students when solving tasks involving the calculation of simple, joint and conditional probabilities in double-entry tables; we consider that this could contribute to the design of learning experiences or didactic sequences that allow overcoming such errors.

Methodology

This research follows a qualitative, descriptive-exploratory methodology (HERNÁNDEZ; FERNÁNDEZ; BAPTISTA, 2010). Using the content analysis technique (KRIPPENDORFF, 1997), the responses provided by students to tasks related to reading and calculating probabilities in double-entry tables are analyzed. Specifically, the mastery or non-mastery of Curcio's reading levels 1 and 2 is identified, as well as the strategies and errors in the calculation of simple, joint and conditional probabilities.

The non-probabilistic sample was made up of 75 high school students (25, 22, 14 and 14 students in 1st, 2nd, 3rd and 4th year, respectively), selected by convenience, whose ages ranged from 14 to 19 years old. In Chile, in the first year of secondary school, double-entry tables and the calculation of probabilities are addressed jointly, that is, the calculation of probabilities is presented in double-entry tables, which implies the reading of this type of statistical representations (MINEDUC, 2016). It should be noted that, at the time of carrying out the research, the students had received instruction on this topic (this information was corroborated with the teachers who teach the mathematics course in first middle school).

http://doi.org/10.14393/ER-v30a2023-32

As an instrument, a questionnaire was designed with five tasks related to a contextualized problem (see Figure 1) dealing with the relationship between two dichotomous variables: age and record of a heart attack. When a person is selected at random, a phenomenon composed of two random experiments is performed: the first with two events: A = " younger than or equal to 55 years" and its complement, not A = "older than 55 years", and the second with two other events: B = "has had a heart attack" and its complement, not B = "has never had a heart attack".

FIGURE 1: the problem presented

Problem. We want to study whether there is a relationship between a person's age and having a heart attack. For this purpose, 200 people have been observed in a medical center. The results are as follows:

	Less than or equal to 55 years of age	Over age 55
Has had a heart attack	20	80
He has never had a heart attack	90	10

1. How many people aged 55 years or younger have never had a heart attack?
   Explain your answer.
2. 2. How many people have never had a heart attack?
   Explain your answer.
3. If we choose one of these people at random, what is the probability that the person has had a heart attack?
   Explain your answer and indicate the operations you perform to answer.
4. If we choose one of these people at random, what is the probability of being older than 55 years and, at the same time, having had a heart attack?
   Explain your answer and indicate the operations you perform to answer.
5. If the selected person is older than 55 years, what is the probability that the person has had a heart attack?
   Explain your answer and indicate the operations you perform to answer.

Source: Adapted from Estrada and Díaz (2006).

The first task involves the literal reading of a given piece of data (level 1, reading the data) represented in the double-entry table. The second refers to the interpretation or integration of the data, by making use of the summation algorithm, to provide a piece of data that is not explicitly represented in the table (level 2, read within the data). The third is related to the calculation of the probability of a simple event (simple probability). The fourth refers to the calculation of a joint probability. Finally, the fifth refers to the calculation of the probability that an event will occur, given that another event has occurred

http://doi.org/10.14393/ER-v30a2023-32

(conditional probability). It should be noted that these types of tasks are similar to those proposed in the textbook for the first year of secondary school, provided free of charge by the Chilean Ministry of Education (GALASSO et al., 2016).

In order to evaluate the relevance of the problem statement and the tasks, the questionnaire was validated by expert judgment and pilot application.

The application of the questionnaire was carried out in a single class session, with approximately 45 minutes for its resolution. The teachers of the courses (1st, 2nd, 3rd and 4th years) collaborated with the application by adopting an observer-type character, not getting involved in the process of solving the tasks.

Once the data was collected, a cyclical analysis process was carried out, with a triangulation of experts, and the responses to each task were categorized. For the responses of tasks 1 and 2, the following categories were chosen: mastery and non-mastery, in case the answer is correct or incorrect; in addition, in task 2, the category partial mastery was considered, when the answer presents the requested data without explicit justification, or when a different data than expected is provided due to calculation errors in the addition algorithm. With respect to tasks 3, 4 and 5, the following categories were chosen: correct, in which are grouped those answers that present the value of the probability requested, together with its justification; partially correct, when the answer presents the value of the probability requested, but without an explicit justification, or provides an answer different from that expected due to calculation errors; and incorrect, when the answer is wrong and presents an error, or a probability is provided under the subjective approach (BATANERO, 2005). It should be noted that in all the tasks the category no response was considered, when the task is not addressed, and the response space is left blank.

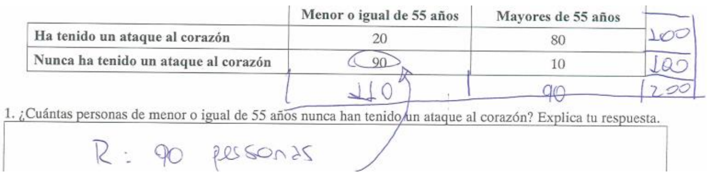
http://doi.org/10.14393/ER-v30a2023-32

Analysis and results

Task 1. Literal reading of a data

In task 1, which requires reading the data, students were expected to provide the requested double absolute frequency, 90 people. Table 2 presents some answers, with their classification.

TABLE 2: Classification of some responses to task 1.

<table>
<tr><th>Category</th><th>Answer / Description (D)</th></tr>
<tr><td>Mastery</td><td>Student 22, first year.


D: Presents the requested double absolute frequency.</td></tr>
<tr><td>Non Mastery</td><td>Student 18, first year.


D: It presents confusion between double absolute frequencies, i.e., confusing n(A ∩ no B) with n(A ∩ B).</td></tr>
</table>

Source: Prepared by the authors.

Table 3 below summarizes the results of the analysis of students' answers to Task 1, by category and year.

TABLE 3: Frequency (and percentage) of mastery of level 1, read data, by category and year.

Category	1°	2°	3°	4°	General
Mastery	20(80)	20(91)	11(78,6)	13(92,9)	64(85,3)
No Mastery	3(12)	2(9)	3(21,4)	1(7,1)	9(12)
No Answer	2(8)				2(2,7)
Total	25(100)	22(100)	14(100)	14(100)	75(100)

Source: Prepared by the authors.

http://doi.org/10.14393/ER-v30a2023-32

The largest proportion of students' answers to task 1 are classified in the mastery category (80%, 91%, 78.6%, and 92.9%, by year, respectively), so this task can be considered as achieved by the students. On the other hand, these results suggest that it is less possible for a student of higher academic level not to answer the task. It should be noted that in all the answers classified in the no mastery category, the students confuse n(A ∩ no B) with n(A ∩ B), being the third year the year with the highest percentage of wrong answers.

Task 2. Reading inside the data

In task 2, which requires the level read within the data, students were expected to provide the requested marginal absolute frequency, 100 people, explaining their answer (e.g., alluding to the sum algorithm "90+10"). Table 4 presents some answers, with their classification.

TABLE 4: Classification of some answers to task 2.

Category	Answer / Description (D)
Mastery	Student 8, second year. 100 personas, se suman las personas que nunca han tenido un ataque al corazón, 90+10. D: Presents the requested marginal absolute frequency together with a formal arithmetic explanation.
Partial mastery	Student 6, third year. 100 Personas D: Presents the requested marginal absolute frequency but does not explain arithmetically and/or with statements its solution.
No mastery	Student 4, fourth year. 10 personas según lo explicito D: Presents confusion between the marginal absolute frequency requested with a double absolute frequency; confusing n(not B) with n(not A ∩ not B).

Source: Prepared by the authors.

http://doi.org/10.14393/ER-v30a2023-32

Table 5 summarizes the results of the analysis of students' answers to Task 2, by category and year.

TABLE 5: Frequency (and percentage) of mastery of level 2, read within data, by category and year.

Category	1°	2°	3°	4°	General
Mastery	14(56)	12(54,5)	8(57,1)	9(64,3)	43(57,4)
Partial mastery	10(40)	10(45,5)	6(42,9)	4(28,6)	30(40)
No mastery				1(7,1)	1(1,3)
No answer	1(4)				1(1,3)
Total	25(100)	22(100)	14(100)	14(100)	75(100)

Source: Prepared by the authors.

The largest proportion of student answers to task 2 are classified in the mastery category (56%, 54.5%, 57.1% and 64.3% per course, respectively). In addition, there is evidence of a high number of answers categorized as partial mastery, in which the requested data is provided, but without presenting arithmetic explanation and/or statements to validate their answer (40%, 45.5%, 42.9% and 28.6% per course, respectively). The above shows that almost all students present a total or partial mastery of reading level 2.

Task 3. Calculation of a simple probability

In task 3, students were expected to calculate the probability of a simple event from Laplace's rule or the calculation of the marginal relative frequency per row, making use of the algorithm (a+b)/(a+b+c+d): P(B) = P(one person had a heart attack) = (20+80)/(20+80+90+10) = 100/200 = 0.5. Table 6 presents some of the answers, with their rankings.

TABLE 6: Classification of some answers to task 3.

<table>
<tr><th>Category</th><th>Answer / Description (D)</th></tr>
<tr><td rowspan="3">Correct</td><td>Student 13, fourth year.
100/200 = 1/2   20+80=100   20+80+90+10=100   Sume los que han tenido un ataque y los dividí en el total.
D: The student provides the correct probability, based on the use of Laplace's rule.</td></tr>
<tr><td>Student 21, third year.
Sería un 50%, ya que esto lo vemos con el ejercicio de la regla de 3.   % 100 | 50   200 | 100   10'000/200=50
D: The student provides the correct probability, using the rule of three. The student calculates the probability from the operation (100*100)/200, considering that 100% corresponds to 200.</td></tr>
<tr><td>Student 5, first year.
la posibilidad es de 50% de probabilidad ya que de 200 100 es la mitad = 50%
D: The student provides the correct probability, from the use of proportionality of quantities, indicating that 100 equals half of the total (200).</td></tr>
<tr><td>Wrong</td><td>Student 1, fourth year.
SI NO / SI: SI, SI NO / NO: SI NO, NO   * Un 25% es la probabilidad que alguna persona haya sufrido un ataque al corazón. * Ocupe esa tabla
D: D: The student presents the equiprobability bias by assigning the same probability to each event linked to the double frequencies.</td></tr>
</table>

Source: Prepared by the authors.

Table 7 summarizes the results of the analysis of student answers to Task 3, by category and course.

TABLE 7: Frequency (and percentage) of answers to task 3, by category and year.

Category		1°	2°	3°	4°	General
Correct	Laplace's Rule	13(52)	9(40,9)	8(57,1)	8(57,1)	38(50,7)
	Rule of three	4(16)	10(45,5)	6(42,9)	2(14,3)	22(29,3)
	Proportionality	3(12)			2(14,3)	5(6,7)
Partially correct	Without explanation	1(4)	2(9,1)		1(7,1)	4(5,3)
Wrong	Equiprobability				1(7,1)	1(1,3)
	Confusion of favorable cases with potential cases		1(4,5)			1(1,3)
	Confusing P(B) with P(not A ∩ B).	1(4)				1(1,3)
No answer		3(12)				3(4)
Total		25(100)	22(100)	14(100)	14(100)	75(100)

Source: Prepared by the authors.

Most of the students' answers to task 3 are classified in the correct category (80%, 86.4%, 100% and 85.7% per course, respectively). In addition, it is identified that: 1) the most used strategy was Laplace's rule in the first, third and fourth year, and the rule of three in the second year; and 2) there were few answers classified as partially correct and incorrect, present in the first, second and fourth years. It should be noted that the presence of the use of proportionality, although minimal, reflects that students are able to calculate probability by means of the relationship between different data in the table, considering the total sample as 100%.

http://doi.org/10.14393/ER-v30a2023-32

Task 4. Calculation of a joint probability

In task 4, students were expected to calculate the probability of the conjunction of two events (joint probability), from Laplace's rule or double relative frequency calculation, making use of the algorithm (b)/(a+b+c+d): P(not A∩B) = P(a person is older than 55 and has had a heart attack) = (80)/(20+80+90+10) = 80/200 = 0.4. Table 8 shows some answers, with their classification.

TABLE 8: Classification of some answers to task 4.

Category	Answer / Description (D)
Wrong	Student 17, fourth year. 80% de posibilidad de 100% D: Presents the value of a conditional probability [P(not A│B)] instead of the requested joint probability [P(not A∩B)]. Student 19, fourth year. $\frac{80}{90}$ La gran mayoría que son mayores de 55 han sufrido un ataque al corazón Sume a los mayores de 55 y luego vi o saqué el dato de los que han sufrido un ataque al corazón D: Calculates a conditional probability [P(B│not A)] instead of the requested joint probability [P(not A∩B)].

Source: Prepared by the authors.

Next, Table 9 summarizes the results of the analysis of the students' answers to Task 4, by category and year.

http://doi.org/10.14393/ER-v30a2023-32

TABLE 9: Frequency (and percentage) of responses to task 4, by category and year

Categoría		1°	2°	3°	4°	General
Correct	Laplace's Rule	11(44)	8(36,4)	8(57,1)	7(50)	34(45,3)
	Rule of three	3(12)	10(45,5)	5(35,7)	3(21,4)	21(28)
Partially correct	Without explanation			1(7,1)		1(1,3)
Wrong	Confusing P(not A∩B) with P(not A|B).				3(21,4)	3(4)
	Confusing P(not A∩B) with P(B|not A).				1(7,1)	1(1,3)
	Confusing P(not A∩ B) with P(A∩no B).	2(8)	1(4,5)			3(4)
	Confusing P(not A∩B) with P(A∩not B)+ P(no A∩B)	1(4)				1(1,3)
	Subjective approach	2(8)	1(4,5)			3(4)
No answer		6(24)	2(9,1)			8(10,7)
Total		25(100)	22(100)	14(100)	14(100)	75(100)

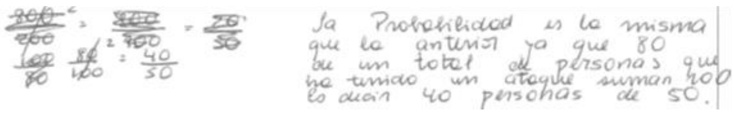
Source: Prepared by the authors.

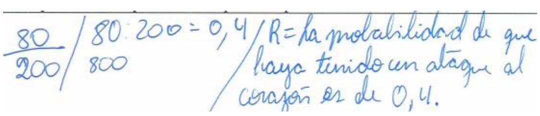
The largest proportion of students' answers to task 4 were classified in the correct category (56%, 81.9%, 92.8% and 71.4%, by course, respectively), with the 3rd year having the highest percentage. The most used strategies were Laplace's rule in the first, third and fourth year (44%, 57.1% and 50%, respectively) and the rule of three in the second year (45.5%). In relation to partially correct answers, only one case was presented in 3rd year, when providing the requested probability, but without an arithmetic explanation and/or statements to validate the answer; and regarding incorrect answers, the highest proportion was presented in 4th year students, when confusing the requested joint probability with the conditional (28.5%). It is worth noting that in this task the number of students who do not answer it increases with respect to the previous one.

Task 5. Calculation of a conditional probability

In task 5, students were expected to calculate the probability of an event occurring given that another event has occurred (conditional probability), from Laplace's rule or the calculation of the conditional relative frequency per column making use of the algorithm (b)/(b+d): P(B│no A) = P(a person has had a heart attack, given that they are older than 55) = (80)/(80+10) = 80/90 = 0.88. Table 10 shows some of the answers to this task.

TABLE 10: Classification of some answers to task 5.

<table>
<tr><th>Category</th><th>Answer / Description (D)</th></tr>
<tr><td>Wrong</td><td>Student 1, second year.


D: The student presents confusion between conditional probabilities (fallacy of the transposed conditional), that is, confusing P(B│not A) with P(not A│B).
Student 15, second year.


D: Presents a joint probability [P(B∩no A)] instead of the requested conditional [P(B│no A)].</td></tr>
</table>

Source: Prepared by the authors.

Next, Table 11 summarizes the results of the analysis of the students' answers to Task 5, by category and year.

TABLE 9: Frequency (and percentage) of responses to task 5, by category and year

Category		1°	2°	3°	4°	General
Correct	Laplace's Rule	10(40)	4(18,2)	7(50)	9(64,3)	30(40)
	Rule of three	2(8)	8(36,4)	4(28,6)	2(14,3)	16(21,3)
Partially correct	Without explanation			1(7,1)		1(1,3)
Wrong	Fallacy of the transposed conditional				3(21,4)	3(4)
	Confuses P(B│Ac) with P(B∩Ac)	2(8)	3(13,6)	2(14,3)	1(7,1)	8(10,7)
	Subjective approach	3(12)	4(18,2)			7(9,3)
No answer		6(24)	2(8)	1(4,5)		1(7,1)
Total		25(100)	22(100)	14(100)	14(100)	75(100)

Source: Prepared by the authors.

It is observed that the highest proportion of students' answers to task 5 are classified in the correct category (48%, 54.6%, 78.6% and 78.6%, by course, respectively), with the 3rd and 4th years having the highest percentage. As in tasks 3 and 4, the strategies most used by students were Laplace's rule in 1st, 3rd and 4th year (40%, 50% and 64.3%, respectively), and the rule of three in 2nd year (36.4%). As for the partially correct answers, there was only one case with no explanation; while, in the incorrect category, two errors are presented with a higher proportion, the confusion of the conditional probability with a joint one (12% and 18.2 %, in 1st and 2nd, respectively) and the confusion of the direction of the conditional probability, that is, confusing P(B│no A) with

http://doi.org/10.14393/ER-v30a2023-32

P(not A│B) (8%, 13.6%, 14.3% and 7.1%, in 1st, 2nd, 3rd and 4th, respectively). Likewise, an increase in the number of students leaving the answer blank is observed in relation to previous tasks.

Conclusions

The results obtained show that middle school students mostly present mastery of Curcio's reading levels 1 (reading the data) and 2 (reading between the data), when performing the literal reading of the double frequency in task 1 and the sum procedure to determine the marginal frequency per row in task 2. It should be noted that, in the latter task, a high percentage of responses categorized as partial mastery is observed, when students do not provide an arithmetic explanation and/or statements to validate their solution, which may be caused by not reading the expression: explain your answer, or by some attitudinal aspect of the student not wanting to justify their solution. Regarding these tasks, our results are similar with those obtained in studies carried out with students, pre-service and in-service teachers (DÍAZ-LEVICOY et al., 2016; GARCÍA-GARCÍA et al., 2019; RODRÍGUEZ; SANDOVAL, 2012) and, in addition, suggest the consideration of justification, regarding task related to Curcio's level 2, as an unnecessary one (40% of answers without justification) despite being a necessary component of statistical literacy.

In relation to the tasks involving the calculation of probabilities, we observed that most students answered correctly using Laplace's rule as the most used strategy in the 1st, 3rd and 4th years, and the rule of three in the 2nd year; this high percentage of correct answers coincides with the results reported in Contreras et al. (2010) and Estrada and Díaz (2006; 2007). The errors with the highest proportion of occurrence in the students' answers were confusing the requested joint probability with a conditional probability or with another joint probability (task 4) and confusing the requested conditional probability [P(B│no A)] with its inverse [P(no A│B)], or with a joint probability (task 5); these coincide with those reported by Estrada and Díaz (2006; 2007) and Contreras et al. (2010). On the other hand, in this research, answers are presented under the subjective approach to probability, this possibly due to the educational level of the students, by giving a value to probability according to their degree of personal belief (BATANERO, 2005) and the calculation of probabilities using proportionality, which suggests that participants can calculate probabilities by alternative methods to Laplace's rule or the use of rule of three, which have a more algorithmic nature. Regarding the percentage of students who do not respond to the probability calculation tasks, and those who provide the requested probability without explanation, these are lower than those reported in similar studies (CONTRERAS et al., 2010; ESTRADA; DÍAZ, 2006; 2007); in addition, students in first year are the only ones who do not respond to the tasks of lower complexity (1 to 3), which can be attributed to the fact that they have not yet developed enough competencies and skills to search for solutions to tasks such as those presented.

The findings reported in this research, regarding the mastery of the first reading levels, as well as the strategies and errors in the calculation of probabilities, provide the basis for the design of learning experiences or didactic sequences where the various mathematical concepts and processes around the double-entry tables are highlighted and, thus, achieve greater understanding by the student; for example, address the various meanings of probability, emphasize the use of classical and frequential approaches to calculate or estimate probabilities from data collected in their environment, among other aspects. On the other hand, the number of students who do not answer the tasks suggests the need to investigate their causes, as possible future research. In general, our results show the need to reinforce the training of high school students (from 1st to 4th year) preferably in the calculation of probabilities, encouraging the development of essential components for probabilistic and statistical literacy.

References

BATANERO, C. Significados de la probabilidad en la educación secundaria. Relime. Revista Latinoamericana de Investigación en Matemática Educativa, v.8, n.3, p.247-263. 2005.

CONTRERAS, J. M. et al. Dificultades de futuros profesores en la lectura y cálculo de probabilidades en tablas de doble entrada. En MORENO, M. et al. (Eds.), Investigación en Educación Matemática, v.XIV, n.1, p.271-280. 2010.

CURCIO, F. R. Developing graph comprehension. Reston, VA: National Council of Teachers of Mathematics [NCTM]. 1989.

DÍAZ, C.; DE LA FUENTE, E. Conflictos semióticos en el cálculo de probabilidades a partir de tablas de doble entrada. Biaix, v.24, n.1, p. 85-91. 2005.

DÍAZ-LEVICOY, D. et al. Lectura de tablas estadísticas por futuras maestras de educación infantil. Educação Matemática Pesquisa, v.18, n.3, p.1099-1115. 2016.

ESTRADA, A.; DÍAZ, C. Computing probabilities from two-way tables. An exploratory study with future teachers. En ROSSMAN, A.; CHANCE, B. (Eds.), SEVENTH INTERNATIONAL CONFERENCE ON TEACHING OF STATISTICS. Proceedings… Salvador de Bahia, Brasil. 2006.

ESTRADA, A.; DÍAZ, C. Errores en el cálculo de probabilidades en tablas de doble entrada en profesores en formación. Uno. Revista de Didáctica de las Matemáticas, n.44, p.48-57. 2007.

FRIEL, S.; CURCIO, F.; BRIGHT, G. Making sense of graphs: critical factors influencing comprehension and instructional implications. Journal for Research in Mathematics Education, v.32, n.2, p.124-158, 2001.

GAL, I. Adult´s statistical literacy: Meaning, components, responsibilities. International Statistical Review, v.70, n.1, p.1-25. 2002.

GAL, I. Towards "probability literacy" for all citizens: building blocks and instructional dilemmas. En JONES, G. (Ed.) Exploring probability in school: Challenges for Teaching and Learning. Boston, MA.: Springer, 2005, p. 39-63.

GAL, I. Understanding statistical literacy: About knowledge of contexts and models. En J. M. CONTRERAS; M. M. GEA; M. M. LÓPEZ-MARTÍN; E. MOLINA-PORTILLO (Eds.), Actas del Tercer Congreso Internacional Virtual de Educación Estadística. Granada, España. 2019.

GALASSO, B., MALDONADO, L., Y MARAMBIO, V. Texto del estudiante. Matemática. Primero Medio. Santiago: Santillana. 2016.

http://doi.org/10.14393/ER-v30a2023-32

GARCÍA-GARCÍA, J. I. et al. Comprensión de una tabla estadística por estudiantes universitarios en México y Chile. Revemat. Revista Eletrônica de Educação Matemática, v.14, p.1-16. 2019.

GEA, M. M. et al. Construcción y lectura de la tabla de doble entrada por profesores de Educación Primaria en formación. Educação Matemática Pesquisa: Revista do Programa de Estudos Pós-Graduados em Educação Matemática, v.22, n.1, p.348-370. 2020.

HERNÁNDEZ, R.; FERNÁNDEZ, C.; BAPTISTA, P. Metodología de la Investigación. México: McGraw Hill. 2010.

KRIPPENDORFF, K. Metodología de análisis de contenido. Teoría y práctica. Barcelona: Paidós Comunicación. 1997.

MINEDUC [MINISTERIO DE EDUCACIÓN DE CHILE]. Matemática. Programa de Estudio. Primero Medio. Santiago: MINEDUC. 2016.

POGGIOLI, L. Estrategias de resolución de problemas. Serie enseñando a aprender. Caracas: Fundación Polar. 1999.

RICO, L. Errores y dificultades en el aprendizaje de las Matemáticas. En KILPATRIK, J.; GÓMEZ, P.; RICO, L. (Eds.). Educación Matemática. Errores y dificultades de los estudiantes. Resolución de problemas. Evaluación. Historia. Bogotá: Grupo Editorial Iberoamérica, 1996. p.69-108.

RODRÍGUEZ, F; SANDOVAL, P. Habilidades de codificación y descodificación de tablas y gráficos estadísticos: un estudio comparativo en profesores y alumnos de pedagogía en enseñanza básica. Avaliação: Revista de la avaliação da Educação Superior, v.17, n.1, p. 207-235. 2012.

RODRIGUEZ-ALVEAL, F.; DIAZ-LEVICOY, D.; VASQUEZ, C. Evaluación de la alfabetización probabilística del profesorado en formación y en activo. Estudios pedagógicos, v.44, n.1, p. 135-156. 2018.

SÁNCHEZ, E. La probabilidad en el programa de estudio de matemáticas de la secundaria en México. Educación matemática, v.21, n.2, p.39-77. 2009.

SCHIELD, M. Statistical literacy survey analysis: Reading graphs and tables of rates and percentages. En Sixth International Conference on Teaching Statistics. Proceedings… 2006.

SOCAS, M. Dificultades, obstáculos y errores en el aprendizaje de las Matemáticas en la Educación Secundaria. En RICO, L. (Ed.). La educación matemática en la enseñanza secundaria. Barcelona: Horsori, 1997. p.125-144.

Received in August 2023.

Approved in October 2023.