Differences of high-school sophomores' performance when solving non-authentic tasks and their authentic versions

David Nexticapan-Cortés[a], Estela Juárez-Ruiz[b], José Antonio Juárez-López[c]

[a]High School Mathematics Teacher, Meritorious Autonomous University of Puebla, Mexico.
[b]Full-Time Professor of Mathematics Education, Meritorious Autonomous University of Puebla, Mexico.
[c]Full-time Professor of Mathematics Education, Meritorious Autonomous University of Puebla, Mexico.

Abstract: This study focuses on students´ performance when solving a set of non-authentic tasks and their authentic versions. Through a quantitative survey-type study, four non-authentic tasks and their authentic versions were implemented in a group of 26 high-school sophomores obtained by a simple random sampling. The results indicated a better performance in solving the authentic tasks than in the non-authentic ones. It is confirmed that students can better solve these tasks, even when a slight change is made in their writing.

Keywords: Authenticity, students' performance, task, word problems.

1. Introduction

Currently, teaching is exclusively directed by the teacher with a focus on the accumulation of knowledge but not on the achievement of meaningful learning. Fragmented knowledge is the result of this type of teaching, leading to limited applicability, relevance, appropriateness, and validity in students' daily life (Ministry of Public Education [SEP], 2017).

Most teachers who see their task as the communication of complete and abstract knowledge tend to adopt an expository style, where their teaching methods are pervaded with definitions and algorithmic and rote procedures. Only at the end and in few cases does a contextualized problem appear as an applied situation of what students have supposedly learned in the classroom (Romero, 2013)

Under these difficulties, as Romero (2013) points out, teachers sometimes pose a problem to their students and then under the same level of difficulty, a set of problems, which leads them to repetitive and rote learning.

The student produces the correct answer to a problem, not because he has understood it from the statement, or because he has understood and solved the problem, or because he has dominated the mathematical object involved, but simply because he has established a similarity to analogous problems; thus, he has done nothing but to replicate a solution made by others for him (D'Amore, Fandiño, Marazzani, and Sarrazy, 2019)

Another current difficulty is that many of the problems proposed in the classroom are not "real" simulations of out-of-school situations, but they "are simply ordinary school math tasks disguised with a figurative context outside of school" (Palm, 2006, p. 42). Due to this lack of realism, such tasks may have a negative impact on students' learning, attitudes, and beliefs. Evidently, this situation is not intended to emulate an entirely real-life situation, but to include real-life objects only as a support for students' thinking about concepts and models (Palm, 2006).

An exploratory pilot study in math textbooks aimed at upper middle level students under the theory of authentic tasks by Palm and Nyström (2009) allowed identifying that a low percentage of the problems analyzed on the subject of the Pythagorean Theorem turned out to be authentic, since most of them do not pose real and significant contextualized situations (Pichardo and Juárez, 2020).

Thus, assessing the authenticity of the problems presented in textbooks is suggested before implementing them in the classroom so that students are brought closer to realistic contexts. Palm and Nyström (2009) emphasize that when students do not work with realistic problems, they cannot perform a careful analysis of the situation presented in the task, but they tend to focus only on the numbers given in the task, as well as on the assumptions that every task has a solution, the solution is achievable, and the answer is only a number with no clear relation between mathematics taught at school and real-life problem solving.

However, Cáceres et al. (2015) pointed out that determining the authenticity of a task is a complex process, since several dimensions must be established; for example, the proximity of the event posed in relation to a high probability of finding it in real life, the adequacy of the question asked to the proposed event, the agreement of the information offered with the question stated, the explicit presence in the figurative context of the purpose for which it must be answered, and the data specificity of the proposed situation.

Therefore, a clinical interview was conducted with a 16-year-old female student of upper middle school level, where she was asked to solve an non-authentic task and its modification to a more authentic one under the

taxonomy of authenticity proposed by Palm and Nyström (2009). The results obtained from the analysis of their mathematical solution processes indicated that when the task was non-authentic, they were very limited and incorrect. Conversely, when the task was modified to a more authentic one, her processes improved substantially because, even with errors, she was able to answer the questions and perform an induction process to generate a mathematical model that consisted of an algebraic expression (Nexticapan & Juárez, 2021).

Consequently, teachers must include authentic tasks or task questionnaires that are authentic or as authentic as possible and validated by a group of scholars so that they can be used as instruments of assessment and reinforced or applied knowledge, and to enhance the teaching and learning process of the students.

On the other hand, Kilpatrick et al. (1998) stressed the fact that the mathematics curriculum design in school in most countries emphasizes the development of reasoning from problem-solving skills. For that reason, problem-solving processes have been an important research area, since it constitutes a powerful teaching tool to develop important skills among students, which is an easily transferrable strategy for life, as it allows the learner to face situations and problems that must be solved (Pérez and Ramírez, 2011).

Considering the difficulties described above, it is necessary to examine the students' performance in upper middle school in a set of non-authentic tasks and their subsequent modification to more authentic tasks that help to identify whether there are substantial changes in their solution processes. Thus, it is appropriate to pose this general research question: What changes become evident in the students' performance when they solve a set of non-authentic tasks and their subsequent modification to more authentic tasks?

This research work is justified by the hypothesis that approaching a problem with increased authenticity, even by only modifying the text related to the task, causes students to increase their tendency to use their real-world prior knowledge effectively in problem solving, according to a study by Palm (2006).

2.Something about authenticity

Archbald and Newmann (1988, cited in Palm, 2006) first introduced the term *authenticity* in learning and assessment as "qualities necessary for many important human achievements beyond success in school" (p. 42). Subsequently, Newmann, Secada, and Wehlage (1995, cited in Palm, 2006) stated that a task "asks students to address a concept, problem, or issue that is similar to one they have encountered or may encounter in life beyond the classroom" (p. 42).

The term authenticity was used by Wiggins (1993) to express his concern for the particular mastery of the several roles and situations that competent professionals encounter in their work. Wiggins states that "the factors involved in authentic evaluation regardless of the subject include tasks that are replicated from or analogous to the types of problems faced by adult citizens and consumers or professionals in the field" (p. 206).

From different interpretations of the term authenticity, Palm (2006) became interested in developing a framework of agreement between mathematical schoolwork and real-world situations beyond the classroom, taking as a reference point the work by Wiggins (1993) to make a further detailed and finer description of what constitutes world-related problems, authentic tasks or tasks in a context with real-life references. According to Palm (2006), this framework can be used to analyze the authenticity of a task, modify it, and make it more authentic.

3.Theoretical framework

As a result, Palm's theory of authentic task situations (2006) refers to the correspondence between verbal problems and real-world situations. This correspondence is based on the simplified representation (simulation) of these situations. Palm (2006) proposed the following aspects of real-life situations that are considered important in his simulation: event, question, information and data, presentation, solution strategies, circumstances, solution requirements, and purpose; some of which have different sub-aspects, as it can be seen in Table 1.

Table 1. *Aspects of real-life situations considered important in Palm's simulation*

A. Event	**F. Circumstances**
B. Question	F1. Availability of external tools
C. Information/Data	F2. Orientation
C1. Existence	F3. Consultation and collaboration
C2. Realism	F4. Discussion opportunities
C3. Specificity	F5. Time
D. Presentation	F6. Consequences of success or failure in

D1. Manner D2. Language usage **E. Solution strategies** E1. Availability E2. Experienced plausibility	solving tasks **G. Solution requirements** **H. Purpose** H1. Purpose in the figurative context H2. Purpose in the social context

According to Palm (2006):

A restriction of comprehensiveness is always necessary. It is not possible to simulate all aspects involved in a situation in real life and consequently it is not possible to simulate out-of-school situations in such a way that the conditions for the solving of the task will be exactly the same in the school situation. (p. 43)

A more detailed description of these aspects drawn from the framework proposed by Palm (2006) is presented below:

A. Event. This aspect refers to the event described in the task. In a simulation of a real-life situation, it is a prerequisite that the event described in the school task has taken place or is highly likely to occur.

B. Question. This aspect refers to the agreement between the task given at the school task and the corresponding out-of-school situation and a corresponding out-of-school task. The question in the school task is one that could actually be posed in the real-life event described.

C. Information/Data. This aspect refers to the information and data of the task and includes given values, models, and conditions. This aspect is divided into the sub-aspects below:

C1. Existence. It refers to the agreement between the information available in the school task and the information available in the simulated situation.

C2. Realism. The realism of values given in schoolwork is an important aspect in simulations of real-life situations, since the students' solution strategies are based on judgments about the reasonability of their responses (Stillman, 1998, cited in Palm, 2006).

C3. Specificity. It refers to the agreement between the specificity of the information available in the school situation and the simulated situation. This pairing is sometimes important for the possibilities of students' reasoning to be similar in the situation inside and outside the school (Baranes, Perry, and Stiegler, 1989, cited in Palm, 2006). For example, the difference between sharing a slice of bread and sharing a cake may cause students to reason differently (Taylor, 1989, cited in Palm and Nyström, 2009).

D. Presentation. This aspect refers to the way in which the task is conveyed to students. This aspect is divided into two sub-aspects:

D1. Manner. It refers to whether the problem is conveyed orally or in writing to students and whether the information is presented in diagrams or tables.

D2. Language usage. It refers to terminology, sentence structure, and amount of text used in the presentation of the task situation. In a simulation of this aspect, with a reasonable degree of reliability, the school task does not include, for instance, difficult terms that prevent students from solving tasks if the corresponding difficulties do not occur in the simulated situation outside the school.

E. Solution Strategies. In a simulation, a task must include the role and purpose of someone solving it. This aspect is divided into two sub-aspects:

E1. Availability. It refers to the agreement between the solution strategies available to students who solve school tasks and those available to people described in the corresponding tasks in real life.

E2. Experienced plausibility. It refers to the agreement between the strategies experienced as plausible to solve the school task and those experienced as plausible in the simulated situation.

F. Circumstances. They are the circumstances under which the task will be solved and are divided into the sub-aspects described below:

F1. Availability of external tools. It refers to tools outside the mind, such as a calculator, a map, or a ruler.

F2. Orientation. It refers to guidance in the form of explicit or implicit suggestions.

F3. Consultation and collaboration. Tasks in real life are solved only individually, through group collaboration, or with the possibility of assistance. In school tasks, these circumstances should be taken into consideration, since the input from other people can affect the type of skills and competencies required to solve those tasks.

F4. Opportunities for discussion. They refer to students asking and discussing the meaning and understanding of the task.

F5. Time. It is important to establish time constraints that do not cause significant differences in the opportunities to solve the task in comparison to the simulated situations.

F6. Consequences of success or failure in the solution of tasks. Different solutions to problems can have different consequences for the solvers. Pressure on the solvers and their motivations to solve the task may affect their solution processes. Encouraging motivation to solve the problem is highly recommended.

G. Solution Requirements. The idea of solution must be interpreted in general terms, that is, both the solution method and the final response to a task. Judgments about the validity of responses and a discussion of the solution methods may be requirements for the solution of the school task.

H. Purpose. This aspect is divided into two sub-aspects:

H1. Purpose in the figurative context. It refers to the fact that in simulations it is essential for the purpose of the task in the figurative context to be as clear to students as it is to the solver in the simulated situation.

H2. Purpose in the social context. It refers to the fact that the school task may be interpreted as the description of a real-life situation, which includes all the properties it contains in real life to be solved. Including a different reasoning strategy and other competencies in the solution of the task are required to fulfill this purpose.

Palm (2006) showed that an increased task authenticity, even when achieved only by modifying the task text, improved the tendency to use their real-world knowledge effectively in solutions to verbal problems.

In later works, Palm and Nyström (2009) used the term *authenticity* to refer to contextualized tasks, which emulate a real-life situation to a reasonable extent. In addition, they pointed out that a school task can never completely simulate an extracurricular situation. However, sometimes the task can be stated and organized in such a way that in many aspects of a real-life situation it is simulated quite satisfactorily, allowing students to solve the task in reasonably close-to-reality conditions.

For Cáceres et al. (2015), determining the authenticity of the tasks is a complex process since several dimensions must be established, such as the proximity of the proposed event in relation to the possibility of finding it in daily life, the suitability of the question asked to the proposed event, the agreement of the information offered with the question posed, the explicit presence of the purpose in the figurative context for which an answer must be offered, and the specificity of the data in the proposed situation.

Palm and Nyström (2009) proposed a brief and synthesized description of only five aspects of the framework presented by Palm (2006) as the most important ones that a task must fulfill to be considered authentic: event, question, information/data, purpose in the contextualized task, and language usage, whose definitions were explained above. In this work, a task that accomplishes these aspects will be considered as an authentic task.

4.Objectives Of The Study

- To examine the students' performance in upper middle school in a set of non-authentic tasks and their subsequent modification to more authentic tasks that help to identify whether there are substantial changes in their solution processes.

5.Hypotheses Of The Study

- This research work is justified by the hypothesis that approaching a problem with increased authenticity, even by only modifying the text related to the task, causes students to increase their tendency to use their real-world prior knowledge effectively in problem solving, according to a study by Palm (2006).

6.Population And Sample

The type of research carried out in this work comes from a quantitative approach in the form of a survey. The population consisted of approximately 120-high-school sophomores from the city of Puebla, Mexico, in the 2020-2021 school year. The sample consisted of 26 students aged between 15 and 17. The sample was obtained by simple random sampling that Cohen, Manion, and Morrison (2012) define as those members of the study

population who are equally probable to be selected and whose probability to be selected will not be affected by the selection of other members of the population; that is, each selection is completely independent from the next. This method involves a random selection from a population list that requires a number of subjects for the sample.

6.1.Statistical Techniques Used in the Present Study

An analysis of the quantitative data obtained was carried out by the Shapiro-Wilk test, which is a statistical test to confirm the normality of the data distribution. According to Arcones and Wang (2006), this test is one of the most reliable tests since it has effective statistical properties, and a sample size of fewer than 50 subjects is needed to conduct it.

6.2.Data Analysis and Interpretation

The data collection instrument consisted of an eight-task questionnaire. Four of the tasks were extracted from the textbook *Mathematics 1 "Algebra in action"* by Joaquín Ruiz Basto, and were reviewed by five scholar judges to assess their authenticity. None of the tasks extracted were authentic, so a modification was made to each one of them to make them authentic from the perspective of Palm and Nyström (2009). The data collection instrument was fulfilled with these four non-authentic tasks and their authentic versions.

Figure 1 presents an example of a non-authentic task that does not fulfill the question, purpose, language usage, and information and data aspects. The scholars agreed that the questions are repetitive, the purpose is not well-defined, the language could be improved, and the rates of each item are very high.

Figure.1. Non-authentic version of the task

Between 3 and 4 million tourists visited a beach vacation resort during the seven days of Holly Week.

The financial expenditure they made was as follows:

Lodging Rates	Food Rates	Tipping Rates
MXN$11,204,105,000	MXN$13,130,102,000	MXN$5,030,000,000

1. What was the average daily expenditure of tourists on each of these areas in a week? And per person? On average, how much did a tourist spend on food per day?
2. On the other hand, one hotel provided 1,200 guests with food that week, but the hotel occupancy rate was 7% lower than expected. Therefore, how many more days did the hotel pantry last?
3. If you were the manager, how would you anticipate the duration of the pantry for any given number of guests?

Figure 2 shows the modification of this task to an authentic one by fulfilling the five aspects of authenticity.

Figure.2. Authentic version of the task

For a graduation trip, the school plans to take a trip to Cancun during the seven days of Holly Week, so you want to know how much money you and your family must save in order to attend to it.

You found on the Internet that between 2 and 3 million tourists visited Cancun during Holly Week last year and that the expenses recorded by all tourists were as follows:

Lodging Rates	Meal Rates	Other Rates
MXN$11,204,105,000	MXN$9,600,585,000	MXN$1,750,000,000

1. On average, how much did a tourist spend on each of these areas in a week? And per day?
2. If the school purchased the round-trip ticket by plane from Mexico City to Cancun at a rate of MXN$1,596 per person, how much money do you think you and your family should save in order to go on the trip?

For data collection, the four non-authentic tasks and their modified authentic versions were implemented to the sampling students in eight 40-minute sessions via Zoom platform, one task at a time. The order of the tasks presented was varied. In some pairs of sessions, the non-authentic task was presented first and then the modified authentic task, whereas the opposite procedure was performed in other pairs of sessions. The purpose of this order was not to bias students on what they would have to solve in each session. Therefore, this study aims to make a quantitative contrast between solving a set of non-authentic tasks and their subsequent modification to more authentic tasks to find out how the students' performance can be enhanced.

At the end of the task, in each session, the students answered a questionnaire of six items related to the aspects of authenticity and then it was submitted digitally following the format in Table 2. This format contains *scale questions*, defined by Corral (2010) as questions whose answers are given through a preset scale, and the grade granted to each student is shown on a Likert scale from 1 to 5, where 1 = *definitely no*, 2 = *probably not*, 3 = *I do not know*, 4 = *probably yes*, and 5 = *definitely yes*.

Table.2. Questionnaire about authenticity aspects

Assertion	1	2	3	4	5
The situation posed in the problem is likely to happen in real life.					
If I experienced the problem situation, I would wonder the same questions posed in the problem.					
The realism of the information and data presented by the problem is identical to those of the same situation that occurs in everyday life.					
The data provided in the problem is sufficient to answer the questions.					
The purpose of the problem is as clear as it would be in real life.					
The problem is written in clear language.					

Analyzing of Data

The written output of each student in the solution of the four non-authentic tasks and the four modified authentic tasks were evaluated using the analytic rubric shown in Table 3, which assesses problem solving, where the grades vary in the range from 50 to 100 possible points.

Table.3. Analytic rubric to evaluate each task

Descr / Grading	Excellent	Weights	Elemental	Weights	Insufficient	Weights
Mathematical Language Management	The mathematical language is used clearly throughout	30%	The mathematical language is used relatively well throughout the	23%	The mathematical language is used poorly throughout the procedure of	15%

	the procedure of the problem to obtain a result.		procedure of the problem to obtain a result.		the problem to obtain a result.	
Formula Management	The appropriate mathematical formulas and expressions are used effectively in all the procedures to solve the problem.	30%	The appropriate mathematical formulas and expressions are used relatively well in all the procedures to solve the problem.	23%	The appropriate mathematical formulas and expressions are used poorly in all the procedures to solve the problem.	15%
Proper *Process-Solution*	The correct result is accomplished through the proper process to solve the problem.	40%	The result is partially accomplished through the proper process to solve the problem.	30%	The correct result is not accomplished to solve the problem.	20%

Note. Adapted from Alsina, Abarca, and Grabulosa (2020).

The reliability of this rubric was assessed by the Cronbach's Alpha coefficient using the grades obtained by the students in the four non-authentic tasks and their modified authentic versions. The Cronbach's Alpha coefficient was 0.802, which is an acceptable value for the reliability of the rubric.

Once each task was evaluated, an Excel file containing all the grades for each student was created and saved to obtain both a grading average when solving the non-authentic tasks and a grading average when solving the modified authentic tasks. Table 4 displays a descriptive summary of these results, where the data indicate that the modified authentic tasks obtained a higher score than the non-authentic ones.

Table.4. A descriptive analysis of the grading averages for the tasks

	NAAvg.	**MAAvg.**
Mean	65	71.65
Median	64.75	69
Mode	67.25	78
Standard deviation	9.99	13.42
Variance	99.82	180.15

Note. NAAvg. = Average obtained by students in the non-authentic tasks. MAAvg. = Average obtained by students in the modified authentic tasks.

On the other hand, Table 5 presents the grading average obtained by each student when solving the non-authentic tasks and the modified authentic tasks.

Table.5. Grading average in the non-authentic task and modified authentic task per student.

NAAvg.	MAAvg.
67.25	95.75
68.75	78.00
65.00	68.25
72.00	66.50
86.50	88.25
54.50	69.00
50.00	56.00
67.00	82.25
50.00	56.00

85.00	89.50
64.50	78.00
58.50	68.00
59.00	65.00
67.50	84.50
60.25	70.75
57.00	69.00
59.00	54.50
52.50	58.00
60.50	50.00
77.25	82.25
84.00	94.00
63.25	54.00
56.50	63.50
67.25	89.75
67.75	61.50
69.25	70.75

Note. NAAvg. = Average obtained by students in the non-authentic tasks. MAAvg. = Average obtained by students in the modified authentic tasks.

Each dataset underwent normality testing, primarily by the Shapiro-Wilk test, which is a test that considers sample sizes of $n \leq 50$. This test states the following hypotheses:

H_0: Data follow a normal distribution

H_a: Data do not follow a normal distribution

The decision criterion for rejecting or not rejecting the null hypothesis can be made through the traditional hypothesis testing method, the p-value method, or the confidence interval method (Triola, 2009)

The normality test was performed with the p-value method, which consists of rejecting the null hypothesis in favor of the alternative hypothesis if $p < \alpha$ (where α is the significance level). Otherwise, if $p \geq \alpha$, the null hypothesis cannot be rejected in favor of the alternative hypothesis.

To verify whether the grading averages obtained by each student in the non-authentic tasks and the modified authentic tasks follow a normal distribution, the Shapiro-Wilk test was used since the sample size is $n = 26$. For this analysis, a 95% confidence level and an $\alpha = 0.05$ significance level was considered. The results are shown in Table 6.

Table.6. Normality test to analyse the grading averages of the non-authentic tasks and the modified authentic tasks

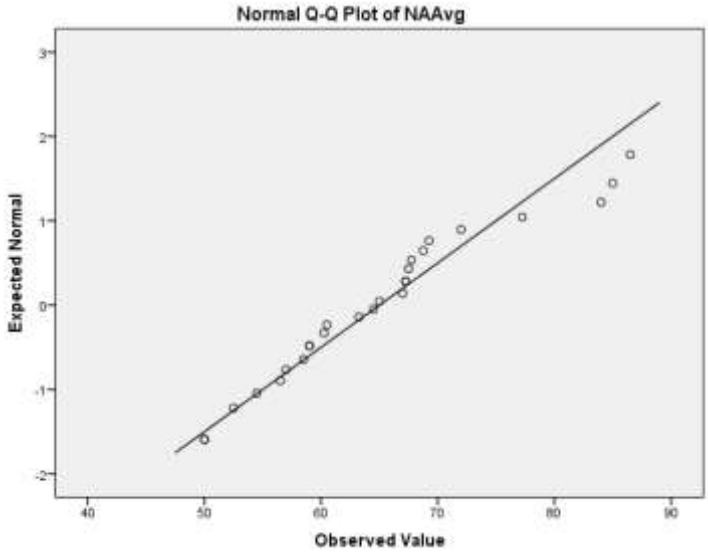
	Statistical Average	**df**	***p*-value**
NAAvg.	.938	26	.118
MAAvg.	.952	26	.262

Note. NAAvg. = Average obtained by students in the non-authentic tasks. MAAvg. = Average obtained by students in the modified authentic tasks.

For the NAAvg. variable, the p-value $p = 0.188$ is greater than 0.05, so H_0 is not rejected in favor of H_a, that is, the grading averages obtained by each student in the non-authentic tasks follow a normal distribution.

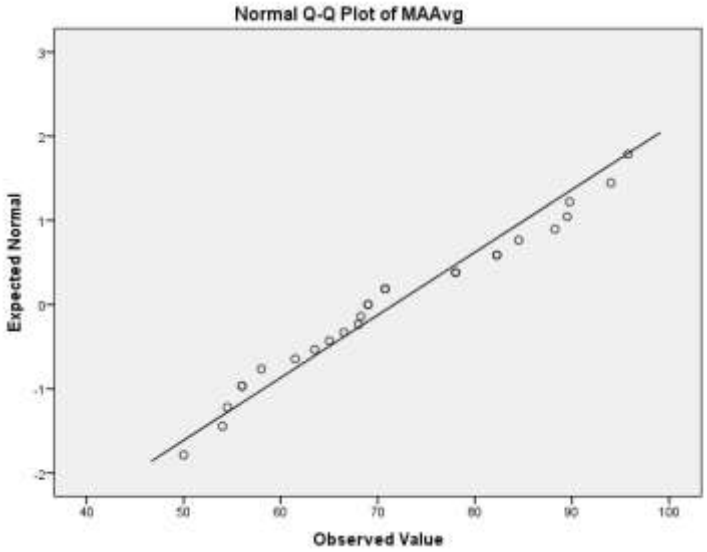
Figure 3 shows the normal quantile plot for the grading averages of these tasks thus, concluding that they come from a normally distributed population since the points are reasonably close to a straight-line pattern with no other systematic pattern.

Figure.3. Normal quantile plot for the grading averages in the non-authentic tasks



Similarly, for the MAAvg. variable, the p-value $p = 0.262$ is greater than 0.05, so H_0 is not rejected in favor of H_a, that is, the grading averages obtained by each student in the modified authentic tasks follow a normal distribution. Figure 4 presents the normal quantile plot for the grading averages of these tasks.

Figure.4. Normal quantile plot for the grading averages in the modified authentic tasks



The grading averages obtained by each student in both the non-authentic tasks and the modified authentic tasks are treated as paired data, since each pair of values represents the same student. To verify whether there is sufficient statistical evidence to conclude that the mean difference in the averages is not zero, inferences from data paired with an $\alpha = 0.05$ significance level are used to prove the assertion that there is a difference between the grading average of the non-authentic tasks and the grading average of the modified authentic tasks.

In Triola (2009), it is pointed out that the following requirements must be fulfilled to make statistical inferences with paired data:

(1) The sample data consist of paired data.
(2) The samples are selected by simple random sampling.

(3) Either or both of these conditions should be fulfilled: the number of paired sample data is large ($n > 30$) or there are differences in the pairs of values that are taken from a population with an approximately normal distribution.

For this study, each of the three requirements is fulfilled:

(1) The sample data consist of paired data concerning the same student. On the one hand, the student obtains the grading average when solving the four non-authentic tasks and the other value is the grading average when solving the four modified authentic tasks.

(2) The sample was selected by simple random sampling.

(3) The number of paired data ($n = 26$) is not large, so the difference normality must be verified. Therefore, the following hypotheses can be stated:

H_0: Differences follow a normal distribution

H_a: Differences do not follow a normal distribution

Table 7 presents the grading average obtained by the students when solving the four non-authentic tasks, the grading average when solving the four modified authentic tasks, and the differences between these pairs of data.

Table.7. Grading averages of the non-authentic tasks and the modified authentic tasks with their differences

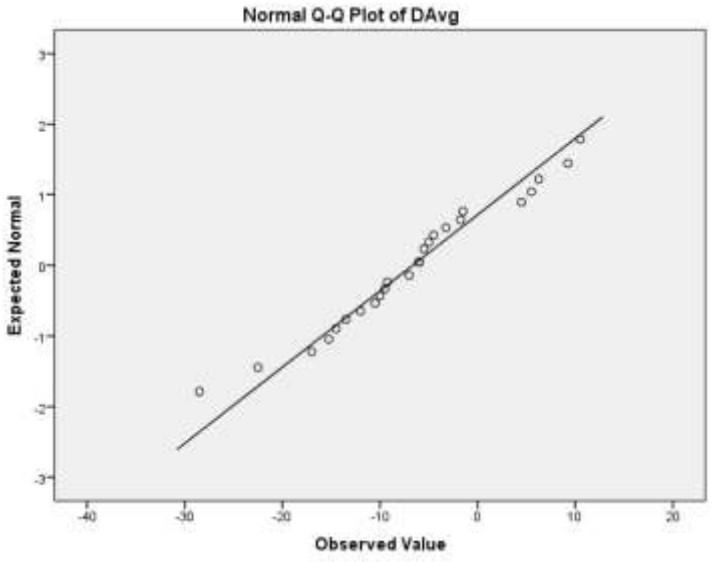
NAAvg.	MAAvg.	DAvg.
67.25	95.75	-28.50
68.75	78.00	-9.25
65.00	68.25	-3.25
72.00	66.50	5.50
86.50	88.25	-1.75
54.50	69.00	-14.50
50.00	56.00	-6.00
67.00	82.25	-15.25
50.00	56.00	-6.00
85.00	89.50	-4.50
64.50	78.00	-13.50
58.50	68.00	-9.50
59.00	65.00	-6.00
67.50	84.50	-17.00
60.25	70.75	-10.50
57.00	69.00	-12.00
59.00	54.50	4.50
52.50	58.00	-5.50
60.50	50.00	10.50
77.25	82.25	-5.00
84.00	94.00	-10.00
63.25	54.00	9.25
56.50	63.50	-7.00
67.25	89.75	-22.50
67.75	61.50	6.25
69.25	70.75	-1.50

Note. NAAvg. = Average obtained by students in the non-authentic tasks. MAAvg. = Average obtained by students in the modified authentic tasks. DAAvg. = differences between these pairs of data.

To determine whether the differences follow a normal distribution, the Shapiro-Wilk test is applied considering a 95% confidence level and an $\alpha = 0.05$ significance level. When applying this test, the results indicate that $p = 0.735 \geq 0.05$, so H_0 is not rejected in favor of H_a, that is, the differences between the pairs of values follow a normal distribution.

Moreover, Figure 5 shows the normal quantile plot of the differences, so it may be concluded that they come from a normally distributed population since the points are reasonably close to a straight-line pattern with no other systematic pattern.

Figure.5. Normal quantile plot for the differences between the averages



A hypothesis test can be performed once the three requirements have been fulfilled.

Step 1: The assertion that there is a difference between the grading average of the non-authentic tasks and the grading average of the modified authentic tasks can be expressed as $\mu_d \neq 0$, where μ_d is the mean value of the differences d for the population of all the paired data.

Step 2: The null hypothesis must express equality, but the alternative hypothesis cannot include equality. Therefore, the hypotheses below may be stated:

$$H_0: \mu_d = 0$$
$$H_a: \mu_d \neq 0$$

Step 3: The significance level is $\alpha = 0.05$.

Step 4: The Student's t-distribution is used.

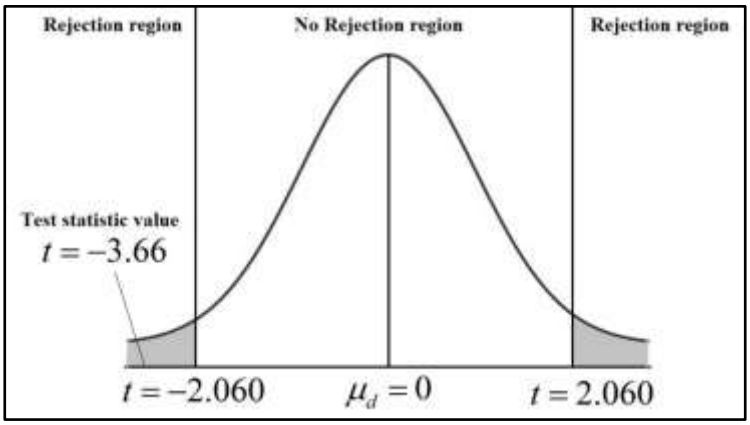
Step 5: The mean value $\bar{d} = -6.65$ and the standard deviation $s_d = 9.26$ of the differences in the paired sample data are calculated.

Step 6: The statistic test value is calculated.

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{-6.65 - 0}{\frac{9.26}{\sqrt{26}}} = -3.66$$

Step 7: The critical values are $t = \pm 2.060$ (Triola, 2009, p. 774). Figure 6 shows the statistic test, the critical values, and the rejected or not rejected region of the null hypothesis.

Figure.6. Rejected or not rejected region of the null hypothesis



Step 8: As the statistic test falls into the critical region, the null hypothesis is rejected in favor of the alternative hypothesis.

The sample data in Table 7 have sufficient statistical evidence to support the assertion that the grading averages of the students when solving the non-authentic tasks and the grading averages when solving the modified authentic tasks are significantly different.

7.Recommendations

- Propose to mathematics teachers the re-design of non-authentic tasks into authentic tasks to test them in the classroom.
- We suggest that mathematics teachers propose in their schools, new ways of analysing and working with verbal mathematical problems or mathematical tasks, which are based on the theory developed by Palm.
- We recommend that teachers and researchers in mathematics education, elaborate as an analysis exercise, the revision and re-writing of problems and tasks in mathematics textbooks that are identified as non-authentic.

8.Conclusion

By responding to the question that guided this research work, results favoured the use of authentic tasks in contrast to non-authentic tasks were yielded. This means that the use of statistical tests proved statistically significant changes in the students' performance when solving authentic tasks in contrast to the non-authentic tasks.

This research work accomplished a first approach to the students' performance when solving authentic tasks. Likewise, the results found in this study indicate that further research needs to be carried out to improve the students' learning process when working with authentic tasks.

A relevant aspect that this study could identify is that the aspects considered in the theory proposed by Palm and Nyström (2009) proved to be useful for the conversion of non-authentic tasks into authentic tasks.

From the results obtained, it was observed that the student's performance improved when they solved authentic tasks in contrast to the non-authentic ones, thus corroborating that a slight change in task formulation, even in writing, causes significant changes in the students' performance.

References

Alsina, Á., Abarca, M., & Grabulosa, I. (2020). Evaluando la competencia matemática: construcción y validación de una rúbrica. *Números: Revista de Didáctica de las Matemáticas, 105*, 119-139.

Arcones, M. A., & Wang, Y. (2006). Some new tests for normality based on U-processes. *Statistics and Probability Letters*, *76*(1), 69–82. https://doi.org/10.1016/j.spl.2005.07.003

Cáceres, M. J., Chamoso, J. M., & Cárdenas, J. A. (2015). Situaciones problemáticas auténticas propuestas por estudiantes para maestro. En C. Fernández, M. Molina & N. Platas (eds.), *Investigación en Educación Matemática X1X* (pp. 201–210). Alicante: SEIEM.

Cohen, L., Manion, L., & Morrison, K. (2012). *Research methods in education*. Routledge. https://doi.org/10.1080/19415257.2011.643130

Corral, Y. (2010). Diseño de cuestionarios para recolección de datos. *Revista Ciencias de la Educación, 20*(36), 152–168.

D´Amore, B., Fandiño, M., Marazzani, I., & Sarrazy, B. (2019). *El contrato didáctico en educación matemática*. Magisterio.

Kilpatrick, J., Gómez, P., & Rico, L. (1998). *Errores y dificultades de los estudiantes Resolución de problemas*

Evaluación Historia. Iberoaméricana.
Nexticapan, D., & Juárez, E. (2021). Análisis de los procesos de resolución de una tarea auténtica y una no auténtica: El caso de Rubí. *Números: Revista de Didáctica de las Matemáticas*, *108*, 27-47.
Palm, T. (2006). Word problems as simulations of real-world situations: a proposed framework. *For the Learning of Mathematics*, *26*(1), 42–47.
Palm, T., & Nyström, P. (2009). Gender Aspects of Sense Making In Word Problem Solving. *Journal of Mathematical Modelling and Application*, *1*(1), 59–76.
Pérez, Y., & Ramírez, R. (2011). Estrategias de enseñanza de la resolución de problemas matemáticos. Fundamentos teóricos y metodológicos. *Revista de Investigación*, *35*(73), 169–193.
Pichardo-Zamora, M. F., & Juárez-Ruiz, E. L. (2020). Análisis de la autenticidad de una tarea desde la perspectiva de un resolutor. *Investigación e Innovación en Matemática Educativa*, 5, 1-25.
Romero, E. (2013). *La ejercitación en el aprendizaje de la matemática* [tesis de licenciatura, Universidad Galileo]. Repositorio Institucional.
SEP. (2017). *Planes De Estudio De Referencia Del Marco Curricular Común de la Educación Media Superior*. http://sems.gob.mx/work/models/sems/Resource/12491/4/images/libro.pdf
Triola, M. F. (2009). *Estadística*. Pearson.
Wiggins, G. (1993). Assessment: authenticity, context and validity. *Phi Delta Kappan*, *75*, 200–214.