On a typology of errors in integral calculus in secondary school related to algebraic and graphical frames.

Anass El Guenyari[1], Mohamed Chergui[2], Bouazza El Wahbi[3]

[1] LAGA Laboratory, Faculty of Science, Ibn Tofail University, Morocco,

[2] EREAM Team, Regional Centre for Education and Training Professions, Kenitra, Morocco,

[3] Department of Mathematics, LAGA Laboratory, Faculty of Science, Ibn Tofail University, Morocco

anass.elguenyari@uit.ac.ma; chergui_m@yahoo.fr; bouazza.elwahbi@uit.ac.ma

Abstract: The present study falls into the efforts to improve practices for addressing errors produced by learners in various situations involving the calculation of integrals. We attempt to clarify as precisely as possible the types of errors that secondary school students produce when using integrals in algebraic and graphical frames. Based on the synthesis of several works dealing with errors specific to integral calculus, we have been able to outline a typology of possible errors that can be produced by students in secondary school. We determine some subcategories for the three known categories of errors: conceptual, procedural, and technical.
After administering a test to a random sample of secondary school students and conducting a principal component analysis, we were able to deduce that in the algebraic frame, certain conceptual and procedural subcategories dominate, with a notable advance for errors due to failure to recognize the integrand function. In the graphical frame, errors related to technical subcategories represent a major source of the erroneous productions of the students tested.

Keywords: Teaching integrals, errors, misconceptions, interplay of frames.

INTRODUCTION

The concept of definite integral, like most concepts in real analysis, is polysemous. It can be interpreted in terms of area, primitives, or the limit of a sum. This diversity is also due to the fact that it is used in several disciplines. For example, it is used to calculate the mean value of a given quantity over a bounded interval. In many countries, it is taught in secondary schools and continues to be taught in higher education. It is part of what is known as modern analysis.
However, several dysfunctions have been pointed out in teaching and learning practices for integrals, as revealed by the authors (El Guenyari, Chergui, & El Wahbi, 2022). They concluded

that the lack of implementation of frame-changing activities and the conversion of registers of semiotic representations negatively impacts the learning of integrals. Here, a frame and a semiotic representation register are used with the same senses as stated by Douady (1986) and Duval (1993) respectively. Based on Donaldson's (1963) classification of errors, Orton (1983) observed difficulties with algebraic and graphical frames in high school and university learners on questions concerning integration and limits. The symbols used to write and calculate integrals were also found to be a source of difficulty.

In an attempt to provide more operationality to the analysis of student integral errors, Seah (2005) conducted research from which he was able to draw the conclusion that students had difficulty with problems involving the integration of trigonometric functions and the use of integration to calculate the areas of specific regions of the plane. When it came to activities dealing with the conceptual elements of integration, students paid less attention than they did with those dealing with the procedural aspects. When asked to determine the area of a surface defined by a function's curve, the x-axis, and two vertical lines, for instance, students failed to take the position of the curve with respect to the x-axis into consideration.

According to the research done by Muzangwa and Chifamba (2012) basic algebra knowledge deficiencies are a significant cause of errors in integral and differential calculus. Additionally, errors and misunderstandings are related to a lack of advanced mathematical thinking, which can be remedied by using a variety of processing frames for mathematical concepts.

The lack of a deeper understanding of the integral was also reported by Ely (2017) after finding that students were unable to cope with situations slightly modified from those with which they were familiar. Ely (2017) explained this vulnerability by pointing out that students had just acquired procedural knowledge of integration in terms of techniques without achieving adequate conceptual knowledge of the underlying structures. As an example, Darvishzadeh et al. (2019) observed that procedural errors are due to confusion between the processes of differentiation and integration.

V. L. Li et al. (2017) deduced, through a study carried out with higher education students, that conceptual errors have many consequences, such as the erroneous use of symbols like dx, the implementation of integration techniques by parts or by change of variables, and the inability to recognize the determination of primitives of usual functions. For this last reason, some students gave the following incorrect answer: $\int -\frac{2}{x}+\frac{3}{x+2}+\frac{2}{(x+2)^2}dx = -2ln|x| + 3\,ln(x+2) + 2ln|(x+2)^2| + c$. In fact, both primitives $ln(x+2)$ and $2ln|(x+2)^2|$ suggested by students are not correct.

Errors related to the manipulation of the bounds of the integral and the variable indicator dx have also been observed by Khanh (2006) with Vietnamese students who cannot understand that one can talk about an interval associated with a primitive, and they consider dx to be a useless factor and omit it from their productions.

It is interesting to note that, in parallel with these attempts to delimit as far as possible the sources of errors produced by learners in calculating integrals, work was also underway to develop the field of didactics dealing with the study of errors and misconceptions. In this context, we refer the interested reader to (Rushton, 2018; Ahuja, 2018), for example. This work covers both didactic and cognitive aspects (Porth, Mattes, & Stahl, 2022). Thus, several error typologies have been developed (Rong & Mononen, 2022) to better understand the nature of these errors and to set up effective remedial processes.

In this work, we are interested in classifying the errors produced by secondary school students in activities involving integrals. This classification has a cognitivist focus. More specifically, we focus on the types of errors made by secondary school students when dealing with integrals in both algebraic and graphical frames. Thus, we attempt to answer the following two main research questions:

- What types of errors are produced by high school students when dealing with the concept of integrals algebraically and graphically?
- What are the main factors that explain the types of errors that can be observed?

CONCEPTUAL FRAME

According to Descomps (1999), an error is a process that marks a difference between the reference point fixed by the didactic contract and the erroneous production. It should be recalled that the didactic contract is defined by criteria set by the teacher, based on the prescriptions of the curriculum and teaching resources. So, by retaining from this reminder that the various actors responsible of the contract are the ones who found the pedagogical practices, we can affirm that the error strongly depends on the context in which it appears. In other words, a statement that may be considered true in one situation may no longer be so in another. Errors are part of learning and provide information for both teacher and learner.

According to Fiard and Auriac (2005), error is essentially the product of a difference between what is produced and what a subject was expected to produce, in view of what he or she was assumed to know how to do. For these two authors, a student's error reflects his procedures, conceptions, or representations that are erroneous and not adapted to the context. However, the student who makes an error is not aware of it because he thinks that he is reasoning adequately.

The advantage of these characterizations of error is that they exclude any moral judgment on students' productions and place the responsibility for error on the student. On the contrary, Fiard and Auriac believe that error is useful for both teacher and learner, as it indicates the mental processes involved in learning.

In the literature, a clear distinction is made between the three concepts of error, difficulty, and obstacle. Difficulty refers to any condition in a situation that increases the probability of producing errors. Language difficulties and disturbances in the development of certain academic skills are examples that may well illustrate the meaning of a difficulty (Chergui, Zraoula, & Amal, 2019).

The obstacle is a witness to the slowness, regressions, and analogies that emerge during the thought formation process (Astolfi, 2015, p. 44). Obstacles encountered in the learning process manifest themselves materially in the production of observable errors. So, the two concepts, errors and obstacles, are complementary. Errors may be due to limitations in the student's intellectual capacities. In this case, we speak of an ontogenetic obstacle. An obstacle is described as epistemological when the knowledge acquired by the student does not enable him or her to carry out a new task proposed by the teacher. The third type of obstacle is called didactic, and includes everything to do with the didactic system put in place by the teacher: poorly formulated instructions, problems relating to the organization of the lesson, interpersonal relations, didactic transposition, and so on.

Given the importance of errors in teaching and learning processes, a number of studies have focused on classifying them. Donaldson (1963, pp. 183-185) identifies three types of error: structural, arbitrary, and executive. The first is due to the inability to appreciate the relationships involved in the problem, the second is due to the student's failure to take account of the constraints established in what is given; and the final one is caused by the inability to perform manipulations while understanding the underlying ideas.

Other studies have used the stages of problem solving established by (Newman, 1977) to identify student errors. These are: reading the statement, comprehension, transformation, process skills, and coding. Based on these procedures, the Australian Ivan Watson (1980) identified eight types of error as follows:

- Inability to read the statement of the situation. For example, he does not recognize words or symbols.
- Inability to understand the situation. This refers to general comprehension. For instance, the meaning of certain terms or symbols.
- Difficulty in identifying the mathematical processes required to obtain a solution.
- Technical difficulties manifest themselves in the inability to perform the mathematical operations required for the task.
- Coding problems are reflected in the inability to write the answer in an acceptable form.
- Motivational problems. The student would have solved the problem correctly if he had tried.
- Errors due to carelessness. These are inattentional errors that are unlikely to be repeated.
- Errors caused by the inappropriate way in which the problem was presented.

It should be noted that categories 2 and 8 are not identical. The first refers to the student who may misunderstand the statement, while the second indicates that the statement or instruction is inadequately formulated.

Based on Donalson's (1963) typology, Orton (1983) conducted a study of students' performance in calculus. Student responses to tasks concerning integration and limits indicated that students had difficulty understanding that integration is the limit of a sum and that there is a relationship between a definite integral and areas under the curve. According to him, many teachers accepted the fact that integration could not be made easy and reacted in various ways. In order to examine students' thinking and misconceptions in dealing with the Riemann integral. An investigation conducted by Thomas and Ye (1996) indicated that students' adherence to an instrumental and procedurally oriented way of thinking, which obstructed them from grasping crucial concepts, resulted in a lack of conceptual knowledge on their part.

In light of the work of Donaldson (1963) and Orton (1983), Seah (2005) has developed a conceptual framework for classifying the various errors and misconceptions that students may encounter when solving integration problems. The errors that students may make have been classified into the following three categories:

- Conceptual errors manifested by the failure to grasp the concepts in the problem or to appreciate the relationships in the problem.

- Procedural errors are attributed to failure to carry out manipulations or algorithms, although concepts in the problem are understood.
- Technical errors are due to a lack of mathematical content knowledge in other topics or to carelessness.

It is in the light of these three complementary aspects that our exploration of the types of errors committed by learners in integral calculus will be undertaken.

METHODOLOGY

To provide answers to the questions posed in this study and with reference to the literature review outlined above, we will use a test to explore the errors made by secondary school pupils. The results obtained will be analyzed both quantitatively and qualitatively.

Data collection

The Riemann integral is part of the course for the final year of secondary school in Morocco (MEN, 2007). The course begins with a presentation of the definition of the Riemann integral over an interval [a,b] using the Newton-Leibniz formula, followed by a statement of the computational properties and the technique of integration by parts. As applications of definite integrals, the program (MEN, 2007) stipulates applications to the calculation of the area of a part of the plane or of a volume.

Our investigation will be undertaken via the test, which is made up of 11 questions, divided into algebraic questions from Q1 to Q6 and graphical ones from Q7 to Q11. A statement of questions and possible answers is presented in Table 1.

Questions	Response strategies
Q.1 Let f be a continuous numerical function on [1, 5] such that $\int_1^5 f(x)\,dx = 10$. F denotes a primitive function of f on [1,5]. Evaluate $F(5) - F(1)$.	Implementing the relation $\int_a^b f(x)dx = F(b) - F(a)$ where F is a primitive of f.
Q.2 f is a continuous function defined on [a, b]. Is there any relationship between the following two integrals: $\int_a^b f(x)dx$ and $\int_b^a f(z)dz$?	Using the notation $\int_a^b f(x)dx = -\int_b^a f(z)dz$.
Q.3 Let f be the numerical function defined on $\mathbb{R}$ by $f(x) = (x-3)^2$. Show that $\int_1^3 f(t)\,dt = \frac{8}{3}$.	Employing Newton-Leibniz formula after determination of a primitive function.
Q.4 Calculate the integral: $I = \int_1^3 (e^x + x\,lnx)dx$.	The use of the integral linearity and an integration by parts.
Q.5 Calculate $\int_1^3 \lvert e^x - e^2\rvert\,dx$.	Application of Charles's relation to remove the absolute value.
Q.6 Calculate $\int_0^1 (e^{2y} + x)\,dy$.	Recognition of the variable to be considered in the integration.

Q.7 What is the sign of the following integral, $J = \int_2^1 (x-3)\ln(x)\,dx$? Justify the answer.

Determining the sign of the integrand function and compare the bounds of the integral.

Q.8 Let f be the numerical function defined on $\mathbb{R}$ by $f(x) = x - 1$ and represented in an orthonormal coordinate system $(o;\vec{\imath};\vec{\jmath})$ by the curve (C_f) in Figure 1. Calculate by two methods $\int_2^4 f(x)dx$.

Method 1: Investing the integrals.
Method 2: recognizing the geometric figure concerned by the area calculation.

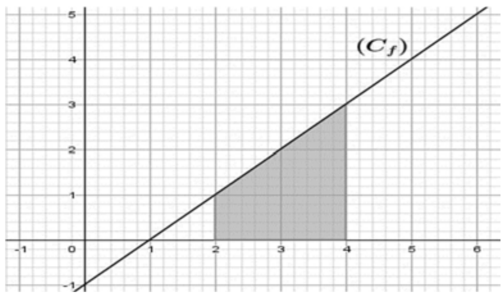


Figure 1

Q.9 Using integrals, express the area A of the domain of the plane colored in gray in Figure 2 below. (C_f) and (C_g) denote the respective curves of two functions f and g.

Using one of the two following formulas:
$A = \int_0^3 f(x)dx + \int_3^4 g(x)dx$ or
$A = \int_0^3 |f(x) - g(x)|dx + \int_1^4 g(x)dx$.

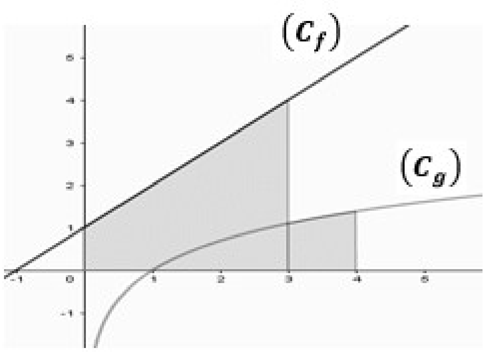


Figure 2

Q.10 Let f be the numerical function defined on $\mathbb{R}$ by $f(x) = 3x^2 - 4$ and (C_f) its representative curve in an orthonormal coordinate system $(o;\vec{\imath};\vec{\jmath})$ (Figure 3). Calculate the area of the part of the plane colored in grey.

Employing the relation $A = \int_{-1}^{1}|f(x)|dx$ and take into account the position of the curve with respect to the abscissa axis.

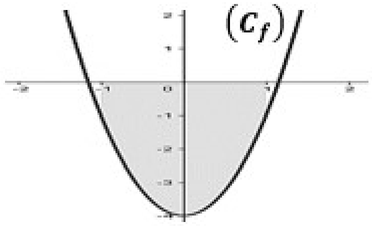


Figure 3

Q.11 Let f and g be two continuous functions on $[0; 2]$ (Figure 4). Express the area of the domain colored in gray by an integral.

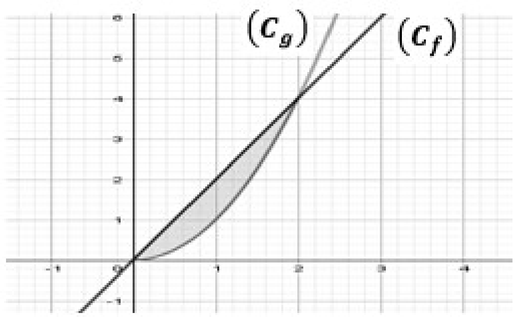


Figure 4

Using the relation $A = \int_0^2 |f(x) - g(x)|dx$.

Table 1 : Test administered

The test was administered in April 2023 to 43 students in their final year of high school (17-18 years old) in the experimental sciences after they had taken the calculus course on integrals. The test participants were from various secondary schools in the Rabat-Salé-Kenitra Regional Education and Training Academy.

Results analysis tools

Based on the conceptual framework developed in the previous section, we set up the grid presented in Table 2, which presents a categorization of integral calculation errors according to algebraic and graphical frames.

Error category	Subcategories related to the algebraic frame	Subcategories related to the graphic frame
Conceptual	Ca1: Combine or confuse primitive and derivation. Ca2: Failure to recognize the integrand function. Ca3: Failure to master the importance of bounds.	Cg1: Lack of understanding of the link between integral and area or volume. Cg2: Inability to recognize the part of the plane concerned by the area calculation.

Procedural	Pa1: Inappropriate choice of operation or property. Pa2: Incorrect implementation of integral calculation by direct determination of a primitive. Pa3: Incorrect implementation of integration by parts.	Pg1: Inappropriate choice of formula for calculating the requested dimension. Pg2: Failure to take into account the position of the curve in relation to the x-axis.
Technical	Ta1: Errors in algebraic calculation Ta2: Errors in applying algebraic properties of common functions. Ta3: Errors in formulating the answer (e.g., forgetting dx, bounds, not placing bounds correctly).	Tg1: Failure to cut out correctly the part of the plan concerned by the surface measurement. Tg2: Failure to read point coordinates correctly.

Tableau 2: Grid for categorizing integral errors according to algebraic and graphical frames

Errors in students' responses to the test questions are classified according to the subcategories shown in Table 2. Each subcategory was encoded to simplify the treatment of the collected data. An enumeration of the numbers in each subcategory is carried out in order to perform an advanced statistical study. For this purpose, we opt for data processing using SPSS software.

RESULTS

Referring to the research questions posed, we will be mainly interested by analyzing the results obtained according to each frame separately, namely the algebraic and graphical frames. The cross-study of these two frames is not the object of this work, nor is it a question of re-exploring the importance of the complementarity between these two frameworks in learning the notion of integrals. But first, let us take a look at a sample of the errors made by the students tested.

Incorrect student productions

After examining the copies of the students tested, we identified the errors listed in Table 4.

Questions	Number of false answers	Errors in learners' productions
Q_1	17	• Since $\int_1^5 f(x)dx = 10$, $f(x) = 10$. So, $F(x) = 10x$, whence $F(5) - F(1) = 40$. • $f(5) - f(1) = 5x - x = 4x$.
Q_2	19	• The answer is yes without giving the relationship. • There is no relationship between the two expressions. • $\int_a^b f(x)dx = [F(x)]_a^b$ and $\int_b^a f(z)dz = [F(z)]_b^a$ (no comparison is given).
Q_3	5	• $\int_1^3 f(t)dt = \frac{1}{2}\int_1^3 (x-3)^2 dt = \frac{1}{2}\ [(x-3)^2]_1^3$.
Q_4	31	• $\int_1^4 xlnxdx = \left[\frac{x^2}{2} lnx\right]_1^4$.

		• Errors in implementing integration by parts.
		• Linearity of the integral not used
Q_5	39	• $\int_1^3 \lvert e^x - e^2 \rvert\, dx = \int_1^3 e^x - e^2\, dx = [e^x - e^2x]_1^3$ • $\int_1^3 \lvert e^x - e^2 \rvert dx = \left\lvert \int_1^3 e^x - e^2 dx \right\rvert$
Q_6	35	• $K = \int_0^1 (e^{2y} + x)\, dy = \left[\frac{e^{2y}}{2}\right]_0^1$ • $K = \int_0^1 (e^{2y} + x)\, dy = \int_0^1 (e^{2y} + y)\, dy$ • $K = \int_0^1 (e^{2y} + x)\, dy = \int_0^1 x dy = \left[\frac{x^2}{2}\right]_0^1$
Q_7	26	• J is negative because $(3 - x)$ is negative on $[1; 2]$.
Q_8	12	• No student was able to calculate the integral by recognizing the figure (trapezoid). • $\int_2^4 x - 1 = \left[\frac{x^2}{2} - x\right]_2^4$ (The absolute value and dx are missing)
Q_9	19	• $A = \int_0^3 f(x)dx + \int_1^4 g(x)dx$ • $S = \int_0^4 f(x) - g(x)\, dx$ • $A = \int_0^3 f(x)dx + \int_3^4 f(x)dx$
Q_{10}	16	• $\int_{-1}^1 3x^2 - 4\, dx = [x^3 - 4x]_{-1}^1$ • $\int_{-1}^1 -f(x)\, dx = [-3x^2 - 4]_{-1}^1 = -8$
Q_{11}	9	• $A = \int_0^1 f(x) - g(x)\, dx.$ • $A = \int_0^2 f(x)dx + \int_0^2 g(x)dx$

Table 3: List of errors made by students

In addition to these errors, numerous algebraic calculation and notation errors were observed. A sample of the students' erroneous productions is provided in the Appendix.

Univariate analysis

Each of the errors listed in Table 3 was classified using the grid in Table 2, taking into account the frame used in the question and the corresponding aspect. To illustrate this, we take the example of the first error in Table 3. The answer given falls within the algebraic frame, and it is clear in this case that the student has not yet acquired that calculating the integral involves a primitive function. Consequently, this error falls into subcategory Ca1.

We used the straightforward descriptive statistics exhibited in Table 4 to analyze the responses in order to get a preliminary overview of the respondent population.

Subcategories	Scores	Mean	Std. Deviation
Ca1	11	,2558	,62079
Ca2	91	2,1163	1,69325
Ca3	26	,6047	,54070
Pa1	86	2,0000	1,19523
Pa2	42	,9767	,59715
Pa3	2	,0465	,21308
Ta1	41	,9535	,89850
Ta2	17	,3953	,54070
Ta3	64	1,4884	1,16235
Cg1	14	,3256	,47414
Cg2	1	,0233	,15250
Pg1	28	,6512	,78327
Pg2	9	,2093	,41163
Tg1	12	,2791	,54883
Tg2	9	,2093	,51446

Table 4 : Descriptive data

Scores were determined by counting the number of occurrences for each subcategory across all student productions. For the averages, indicated in Table 4, they are calculated by considering the total number of students tested. Thus, the first value 0.2558 represents the average of the subcategory Ca1 in the sample studied.

It is interesting to note similarities in certain averages. This concerns pairs of subcategories (Ca2, Pa1), (Pa2, Ta1), and (Pg2, Tg2). To confirm or refute this point, it is convenient to carry out a test of the averages using the t-test of two independent samples. If the p-value is less than the significance level ($p < 0.05$), the difference does not equal zero.

		Paired Samples Correlations		Paired Differences					t	df	Sig. (2- tailed)
		Correlation	Sig.	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper			
Pair 1	Ca2 - Pa1	,424	,005	,1162	1,606	,2450	-,3781	,6107	,475	42	,638
Pair 2	Pa2 - Ta1	-,046	,767	,0232	1,101	,1680	-,3158	,3623	,138	42	,891
Pair 3	Pg2 -Tg2	,126	,423	,0000	,6172	,0941	-,18995	,1899	,000	42	1

Table 5 : Paired Samples Test

We observe that the variables Ca2 and Pa1 are moderately and positively correlated (r = 0,424, p = 0,005). This situation is no longer statistically true for the pairs (Pa2, Ta1) and (Pg2, Tg2) since the significance level exceeds the accepted value. Furthermore, there is no significant average difference between the three evoked pairs. This statement is also confirmed by the fact that the mean of the differences for each pair of variables lies within the confidence interval.

Bivariate analysis

The cross-tabulation of the variables indicating the different subcategories of errors that were identified during the processing of the activities on integrals enabled us to highlight some significant correlations at the 0.05 level (2-tailed), as shown in Table 6.

		Ca1	Ca2	Ca3	Pa1	Pa2	Pa3	Ta1	Ta2	Ta3	Cg1	Cg2	Pg1	Pg2	Tg1	Tg2
Ca1	Correlation	1														
	Sig.															
Ca2	Correlation	,537	1													
	Sig.	,000														
Ca3	Correlation	,308	,285	1												
	Sig.	,044	,064													
Pa1	Correlation	,385	,424	,147	1											
	Sig.	,011	,005	,346												
Pa2	Correlation	,402	,450	,266	,600	1										
	Sig.	,008	,002	,085	,000											
Pa3	Correlation	,088	-,279	-,250	,187	,009	1									
	Sig.	,575	,070	,106	,230	,956										
Ta1	Correlation	-,106	,113	-,039	,089	-,046	-,113	1								
	Sig.	,498	,470	,805	,572	,767	,471									
Ta2	Correlation	,330	,261	-,023	,074	,029	,043	-,010	1							
	Sig.	,031	,091	,885	,639	,853	,783	,948								
Ta3	Correlation	-,078	,140	-,064	,171	-,018	,002	,045	,064	1						
	Sig.	,618	,371	,682	,272	,911	,989	,774	,682							
Cg1	Correlation	,277	,337	,050	,126	,196	,082	-,019	,229	,223	1					
	Sig.	,073	,027	,752	,421	,209	,600	,901	,140	,151						
Cg2	Correlation	-,064	-,011	,114	-,131	,006	-,034	,356	,175	-,200	,222	1				
	Sig.	,682	,946	,466	,404	,969	,828	,019	,263	,199	,152					
Pg1	Correlation	,090	-,023	,116	,229	,033	,100	,010	,109	-,122	,441	,269	1			
	Sig.	,566	,886	,457	,140	,833	,525	,948	,489	,435	,003	,081				
Pg2	Correlation	-,121	-,207	,274	,145	,117	,158	,027	-,060	,080	,253	,300	,601	1		
	Sig.	,438	,184	,076	,353	,454	,312	,864	,704	,611	,102	,051	,000			
Tg1	Correlation	-,145	-,318	-,341	,073	-,198	,294	,172	-,060	,080	,374	,205	,620	,368	1	
	Sig.	,355	,038	,025	,644	,204	,056	,271	,704	,611	,013	,187	,000	,015		
Tg2	Correlation	-,023	,081	,048	,039	-,216	-,091	,279	,123	,024	,104	-,064	,422	,126	,378	1
	Sig.	,886	,607	,761	,805	,164	,562	,070	,430	,878	,505	,686	,005	,423	,012	

Table 6: Pearson correlations between subcategories of errors

In light of these results, which highlight a number of correlations between the variables studied, we feel it would be interesting to make further progress in processing the results obtained. To this end, we will conduct a principal component analysis (PCA).

Principal component analysis

In order to highlight the different subcategories of errors made by the students in their answers to the test on integrals, we carried out a PCA, which allows multivariate analysis of all the variables. PCA is administered with quantitative variables or with measured hierarchical variables. The principle of PCA is to minimize the number of variables. The new variables are called factors and represent linear functions of the initial variables.
The adequacy of the sample must be examined first in PCA (Johnson & Wichern, 2002). To achieve this, two tests can be administered: the Kaiser-Meyer-Olkin (KMO) test and Bartlett's sphericity test. The first gives a proportion of the variance between variables that could be a common variance. It is scored from zero to one, with zero being inappropriate and a value close to one being appropriate. For the Bartlett test, the observed correlation matrix is compared with the identity matrix. In general, KMO values of at least 0.50 and $p < 0.05$ for the Bartlett sphericity test are considered acceptable.

		Graphic frame	Algebraic frame
Number of items		6	9
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		,714	,611
Bartlett's Test of Sphericity	Approx. Chi-Square	62,652	76,061
	df	15	36
	Sig.	,000	,000

Table 7: KMO and Bartlett's Test

From the values obtained, we can deduce that:
- Since the KMO index is sufficiently greater than 0.5, all items are factorable (2006);
- Bartlett's test revealed that the calculated p-value is below the 0.05 level of significance. It is therefore appropriate to reject the hypothesis that there is no correlation significantly different from 0 between the variables and to accept the fact that there are correlations that are not all equal to zero.

With regard to reliability, the Cronbach's coefficient was calculated for the items relating to each frame. The results are as follows:

Frame	Cronbach's Alpha	N of Items
Algebraic items	,721	9
Graphic items	,735	6
All items	,713	15

Table 8: Reliability Statistics

The reliability of our grid is therefore satisfactory. We can therefore conclude that all the items contribute to the reliability of the grid and that no purification is necessary.
To understand student performance in each frame, we carried out a PCA according to each frame.

- **PCA according to graphical frame**

By implementing the PCA on all the items in the graphical frame without previously fixing the number of factors requested, we obtained the results presented in Table 9.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
FG 1	2,668	44,459	44,459	2,668	44,459	44,459
FG 2	1,143	19,058	63,516	1,143	19,058	63,516
FG 3	,774	12,902	76,418			
FG 4	,663	11,042	87,460			
FG 5	,488	8,126	95,587			
FG 6	,265	4,413	100			

Table 9: Total variance explained by applying PCA for graphic frame

Using the Kaiser criterion, the components to be retained are those with an eigenvalue greater than 1. Consequently, the first two components explain more than 63% of the total variance, making a total of 44.45% for the first and 19.05% for the second. The sum of the corresponding eigenvalues is 3,8. This means that these two components can replace almost four items. Note also that the sum of the eigenvalues is equal to 6, which is the total number of items considered. The contribution of each subcategory of errors in forming the principal components is explicated in Table 10.

	FG1	FG2	FG3	FG4	FG5	FG6
Cg1	13,673	3,021	69,373	1,302	9,706	2,926
Cg2	6,855	41,098	3,979	46,931	1,095	0,041
Pg1	29,787	0,680	0,937	2,056	0,265	66,275
Pg2	18,259	5,451	22,565	31,578	3,629	18,519
Tg1	22,898	3,385	0,855	1,729	64,659	6,473
Tg2	8,528	46,365	2,291	16,404	20,647	5,766

Table 10: Contributions of graphic variables (%)

The Component plot of factors 1 and 2 on the F1 (component 1) and F2 (component 2) axes is shown in Figure 5. It corresponds to a projection of the initial variables onto a two-dimensional plane constituted by the two factors.

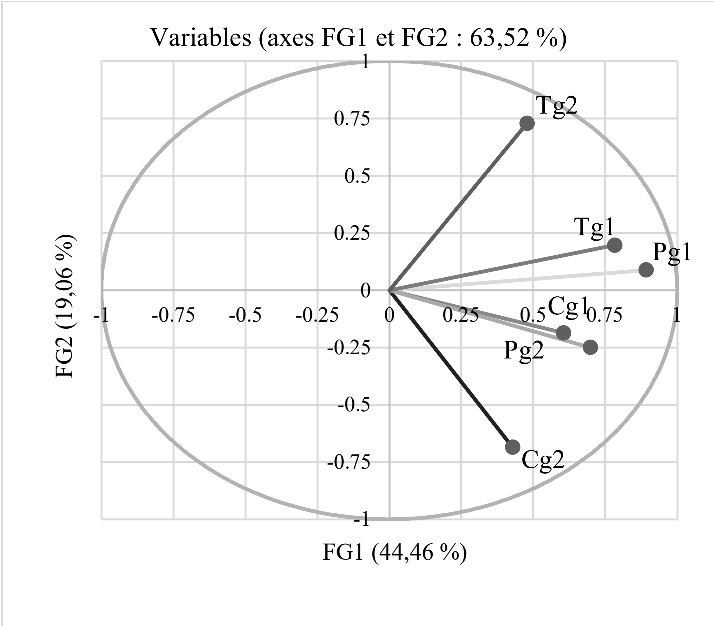


Figure 5 : Circle of correlations for graphic frame

- **PCA according to algebraic frame**

Taking into account the same considerations as in the previous case, we carried out a PCA, which gave the results listed in Table 11.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
FA1	2,636	29,292	29,292	2,636	29,292	29,292
FA 2	1,349	14,987	44,279	1,349	14,987	44,279
FA 3	1,201	13,339	57,618	1,201	13,339	57,618
FA 4	1,113	12,368	69,986	1,113	12,368	69,986
FA 5	,928	10,312	80,298			
FA 6	,662	7,355	87,653			
FA 7	,497	5,520	93,173			
FA 8	,327	3,635	96,808			
FA 9	,287	3,192	100,000			

Table 11: Total variance explained by applying PCA for algebraic frame

We note that four components have eigenvalues greater than 1. There are therefore four components that can be extracted from our grid, and the cumulative variance that they can explain is 70% of the total variance. The contribution of each subcategory of errors to forming the principal components is explained in Table 12.

	FA1	FA2	FA3	FA4	FA5	FA6	FA7	FA8	FA9
Ca1	21,900	1,383	5,794	6,277	0,903	2,256	30,608	1,180	29,698
Ca2	24,139	4,463	3,021	2,491	0,019	5,544	13,952	0,026	46,344
Ca3	8,448	18,189	9,269	1,543	2,906	51,296	6,179	0,087	2,084
Pa1	19,670	8,260	3,983	11,804	0,219	0,146	3,289	50,890	1,739
Pa2	21,379	0,829	0,655	13,185	0,019	10,453	15,618	37,210	0,653
Pa3	0,206	57,374	0,762	1,130	2,225	17,154	1,234	3,544	16,371
Ta1	0,012	5,368	33,851	1,166	47,665	8,010	0,030	2,959	0,938
Ta2	3,912	2,911	0,725	62,303	1,298	0,104	28,670	0,062	0,015
Ta3	0,333	1,223	41,941	0,101	44,746	5,037	0,421	4,042	2,158

Table 12: Contributions of variables (%)

DISCUSSION

From Table 4, we can clearly see that the errors that fall under the subcategories Ca2, Pa1, and Ta3 are the most frequent when dealing with integrals in the algebraic frame. Referring to Table 6, we note that the two subcategories failure to recognize the integrand function and inappropriate choice of operation or property are moderately and positively correlated with a fairly acceptable significance level (p-value = 0.005). But nothing can be confirmed with regard to the correlation between the subcategory Ca2 and errors in formulating the answer. This means that, for the student tested, not recognizing the integrand function has an impact on the choice of operations required to calculate integrals but not necessarily on the ability to formulate answers. This result is very interesting didactically. In fact, the calculation of integrals, whether directly by the Newton-Leibniz formula or by another technique, requires the determination of primitive functions. It is to this latter task that the teacher must then pay attention to mitigate the impact of the inability to recognize the functions to be integrated.

For the graphical frame, procedural errors are dominated by inappropriate choices of formula for calculating the requested dimension, followed by conceptual errors concerning a lack of understanding the link between integral and area or volume. Moreover, the correlation between these last two subcategories (Cg1 and Pg1) of errors is positively medium. This result seems quite logical to us, given that it is unlikely that a student who fails to understand the link between the integral and the geometric quantity to be measured will correctly choose the formula to use.

Note that these results are in harmony with those deduced by Seah (2005) in his study, where he observed difficulties in calculating integrals of trigonometric functions and in their applications in area calculations. We can also state that these preliminary results are aligned with those of Muzangwa and Chifamba (2012). This has motivated us to go further in our analysis of the results obtained, with the aim of better identifying the essential factors that explain the production of errors by students.

Within the same frame, several pairs of error subcategories are highly positively correlated. These include the couples of subcategories (Pa1, Pa2), (Pg1, Pg2), and (Pg1, Tg1). It is interesting to pay attention to the fact that these significant correlations relate to procedural issues in the majority of cases.

Other subcategory pairs are moderately positively correlated. For example, (Ca2, Pa1), (Ca2, Pa2), and (Pg1, Tg2). This last positive correlation between an algebraic subcategory and a graphical one is cognitively meaningful from the study performed by the authors (El Guenyari, Chergui, & El Wahbi, 2022). That is to say, the integral should be invested in various frames in the learning situations provided to students for good cognitive functioning when processing it.

However, it should also be noted that the subcategory Tg1 is negatively correlated with Ca2 and Ca3. This means that the inability to correctly cut out a part of the plane to calculate its area is negatively correlated with incompetence in recognizing the integrand function and with a lack of appreciation of the importance of the bounds of the integral. This result, which does not seem at all normal, questions the conditions for learning the integral among the students tested. More explicitly, why do students not succeed in practicing the change of frames easily?

In addition to this surprising result, which calls for greater precision, there is a lack of information on the correlation between several pairs of subcategories, as reported in Table 6. Remarkably, no statistical results were obtained concerning the correlation between the inability to recognize the part of the plane concerned by the area calculation and all the other error subcategories of the graphical frame. This situation also extends to the algebraic framework by observing, for example, the two subcategories failure to master the importance of bounds and errors in formulating the answers.

In order to clarify these points, we carried out a statistical analysis using PCA. Focusing on axis 1 (FG1) in Figure 5, we see that the first factor FG1 is positively correlated with all of the initial graphical subcategories. This correlation is quite strong with the subcategories Pg1 and Tg1. According to the results in Table 10, these two variables are the most important in forming the principal component FG1. This can be interpreted by the fact that the errors that fall into these two subcategories, which can be considered as forming the FG1 factor, are the main contributors to total variability.

The presence of an acute angle between two variables indicates that they are fairly well correlated. This is the case between several subcategories, as shown in Figure 5. But when the angle is almost right, the variables are rather uncorrelated. The inability to recognize the part of the plane concerned by the area calculation and the failure to read point coordinates correctly fall into this latter case.

With regard to the second principal component FG2, the subcategories Cg2 and Tg2 present a high correlation, which can be clearly visualized by the projection on the vertical axis in Figure 5. To identify this second principal component, it should be noted that the subcategories Cg2 and Tg2 make a major contribution to its formation. In Table 10, their contributions are 41.098% and 46.365%, respectively.

The analysis of the circle of correlations in Figure 5 shows that axis 2 highlights an opposition between subcategories with positive correlations (Pg1,Tg1,Tg2) and those with negative correlations (Cg1, Cg2, Pg2).

From the PCA carried out on the nine subcategories of errors on integrals related to the algebraic frame, it turned out, as shown in Table 10, that four principal components can be extracted. The first, which contributes to explaining over 29% of the variance, is mainly formed, according to Table 12, by the subcategories Ca1, Ca2, and Pa2. The second main component, which explains around 15% of the total variance, is made up mainly of the Pa3 and Ca3 subcategories. The component FA3 is formed by the subcategories Ta1 and Ta3. Finally, the last principal component, FA4, is essentially formed by the subcategory Ta2, with a proportion that exceeds 62%. It is important to note that the principal components FA3 and FA4 are formed by technical subcategories and together contribute in explaining almost 25% of the total variability. So, almost 75% of the types of errors committed by students in the algebraic frame are mainly due to conceptual reasons, followed by procedural ones. Errors due to an inappropriate choice of operation or property are not included in the latter type. The impact of a lack of conceptual knowledge about integrals was underlined by Ely (2017).

CONCLUSION

No one doubts the importance of learning integral calculus in secondary school. It is hard to conceive of a curriculum in which this notion is absent, given its usefulness in other disciplines. However, teaching it, and therefore learning it, poses problems that have a negative impact on the acquisition of other mathematical topics and other skills. Hence the need to find effective ways of attenuating the impact of these problems. One approach is to make the practice of dealing with errors produced by learners in various situations involving the calculation of integrals as reflexive as possible.

The present work fits into this context by attempting to elucidate as accurately as possible the types of errors that secondary school students make when using integrals in algebraic and graphical frames. The literature review we have carried out has enabled us to draw up a typology of errors specific to integral calculus. This typology is the synthesis of several works that have addressed the same theme. We were able to determine subcategories for the three categories of error: conceptual, procedural and technical.

On the basis of this typology, we examined the work of 43 secondary school students on a test involving algebraic and graphical questions. The examination consisted in classifying the errors identified according to 15 possible subcategories.

For the algebraic frame, conceptual errors were dominated by failure to recognize the integrand function, while procedural errors were caused by inappropriate choice of operation or property and incorrect implementation of integral calculation by direct determination of a primitive. The main technical errors are attributable to faulty algebraic calculations or incorrect formulations of the answer.

Principal component analysis showed that all the subcategories relating to conceptual and procedural aspects represent the essential factors responsible for the variability observed in the test results. Several of these subcategories are positively correlated. We cite the example of errors due to not recognizing the integrand function with those arising from the choice of operations required to calculate integrals. This conclusion has an important pedagogical character in teaching practice. It calls for sufficient attention to be paid to identifying the functions to be integrated. This is also justified from an epistemological point of view, as it is well known that for several classes of numerical functions, there are appropriate techniques for determining primitive functions. It is also interesting to note that, in the algebraic context, technical errors do not play a significant role in students' erroneous productions.

For the graphical frame, it was concluded that errors attributed to a lack of understanding of the link between integral and area or volume, an inappropriate choice of formula for calculating the requested dimension, and failure to cut out correctly the part of the plane concerned by the surface are dominant. To overcome this conceptual problem, it is recommended to adopt teaching methods based on the employment of contextualized situations, given their effectiveness in acquiring the meaning of mathematical concepts, as deduced (Sijmkens, Scheerlinck, De Cock, & Deprez, 2022) or in (Naamaoui, Chergui, & El Wahbi, 2023). In addition, the PCA allowed us to extract two main factors that explain student errors. the technical subcategories contribute to the formation of these two main factors. In this regard, it should be mentioned that technical tasks in the graphic frame are not to be undervalued. In fact, as well as being tasks requiring meticulousness, they also demand well-developed cognitive and visual levels.

Acknowledgments

The authors would like to thank the anonymous referees for their insightful comments and suggestions, which significantly improved the quality of this manuscript.

References

[1] Ahuja, A. (2018). Errors as Learning Opportunities: Cases from Mathematics Teaching Learning. In V. Kapur, & S. Ghose, *Dynamic Learning Spaces in Education.* Springer. https://doi.org/10.1007/978-981-10-8521-5_7

[2] Astolfi, J.-P. (2015). L'erreur, un outil pour enseigner (12e éd.). *Issy-les-Moulineaux : ESF Editeur*.

[3] Chergui, M., Zraoula, L., & Amal, H. (2019). Impact des difficultés langagières sur l'apprentissage des nombres complexes. *Actes de la deuxième édition du colloque international sur la formation et l'enseignement des mathématiques et des sciences (CIFEM2018)*, 130-141. El Jadida, Maroc. https://cifem2018.sciencesconf.org/data/pages/Didactique_TIC_Innovation_pedagogique.pdf

[4] Darvishzadeh, M., Shahvarani Semnani, A., Alamolhodaei, H., & Behzadi, H. (2019). Analysis of Students' Mistakes in Solving Integrals to Minimize their Mistakes. *Control and Optimization in Applied Mathematics, 4*(2), pp. 49-60. https://doi.org/10.30473/coam.2020.55039.1148

[5] Descomps, D. (1999). La dynamique de l'erreur. Paris : Hachette livre.

[6] Donaldson, M. (1963). A Study of Children's Thinking. *Tavistock Publications*. London.

[7] Douady, R. (1986). Jeux de cadres et dialectique outil/objet dans l'enseignement des mathématiques. *Revue RDM, 7(2)*, 5-32. https://revue-rdm.com/1986/jeux-de-cadres-et-dialectique/

[8] Duval, R. (1993). Registres de représentation sémiotique et fonctionnement congnitif de la pensée. *Annales de didactique et des sciences congnitives, 5*, 37-65. https://publimath.univ-irem.fr/numerisation/ST/IST93004/IST93004.pdf

[9] El Guenyari, A., Chergui, M., & El Wahbi, B. (2022). A study on the effectiveness of some cognitive activities in teaching integrals in secondary school. *Mathematics Teaching-Research Journal, 14*(5), pp. 66-83. https://commons.hostos.cuny.edu/mtrj/volume-15-n-1/

[10] Ely, R. (2017). Definite integral registers using infinitesimals. *The Journal of Mathematical Behavior, 48*, pp. 152–167. https://doi.org/10.1016/j.jmathb.2017.10.002

[11] Fiard, J, & Auriac, E. (2005). L'erreur à l'école, petite didactique de l'erreur scolaire. Paris : Le Harmattan.

[12] Hair Jr, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). Multivariate Data Analysis. *6ª ed., Upper Saddle River: Pearson Prentice Hall*.

[13] Johnson, R. A., & Wichern, D. W. (2002). Applied Multivariate Statistical Analysis. *Pearson Prentice Hall*, 430. New Jersey.

[14] Khanh, T. L. (2006). La notion d'intégrale dans l'enseignement des mathématiques au lycée : une étude comparative entre la France et le Vietnam. *Thèse*. Université Joseph-Fourier-Grenoble I. https://theses.hal.science/tel-00122062

[15] Li, V. L., Julaihi, N. H., & Eng, T. H. (2017). Misconceptions and Errors in Learning Integral Calculus. *Asian Journal of University Education, 13*(2), pp. 17-39. https://eric.ed.gov/?id=EJ1207815

[16] MEN. (2007). Orientations pédagogiques générales et programme des mathématiques pour le cycle secondaire qualifiant. *Direction des Curricula, Ministère de l'éducation nationale, Royaume du Maroc*.

[17] Muzangwa, J., & Chifamba, P. (2012). Analysis of error and misconceptions in the learning of calculus by undergraduate students. *Acta Didactica Napocensia, 5*(2), pp. 1-10. https://files.eric.ed.gov/fulltext/EJ1054301.pdf

[18] Naamaoui, Y., Chergui, M., & El Wahbi, B. (2023). Attitudes of mathematics and physics teachers towards difficulties in modeling with differential equations in secondary school. *International Journal on "Technical and Physical Problems of Engineering" (IJTPE), 15*(4). http://iotpe.com/IJTPE/IJTPE-2023/IJTPE-2023.html

[19] Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematical tasks. *Victorian Institute for Educational Research Bulletin, 39*, pp. 31-43.

[20] Orton, A. (1983). Students' understanding of integration. *Educational Studies in Mathematics, 14*, pp. 1-18. https://doi.org/10.1007/BF00704699

[21] Porth, E., Mattes, A., & Stahl, J. (2022). The influence of error detection and error significance on neural and behavioral correlates of error processing in a complex choice task. *Cognitive, Affective, & Behavioral Neuroscience, 22*, pp. 1231–1249. https://doi.org/10.3758/s13415-022-01028-6

[22] Rong, L., & Mononen, R. (2022). Error analysis of students with mathematics learning difficulties in Tibet. *Asian Journal for Mathematics Education, 1*(1), pp. 52–65. https://doi.org/10.1177/27527263221089357

[23] Rushton, S. J. (2018). Teaching and learning mathematics through error analysis. *Fields Mathematics Education Journal, 3*, pp. 1-12. https://doi.org/10.1186/s40928-018-0009-y

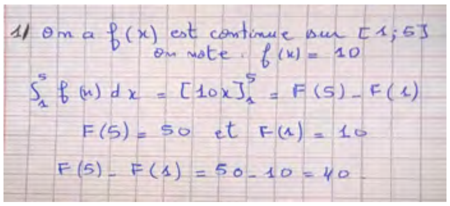
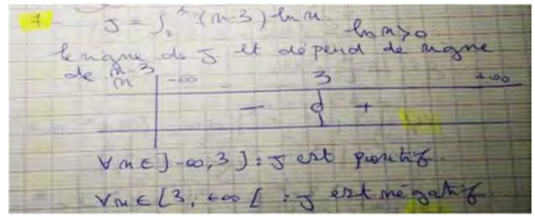
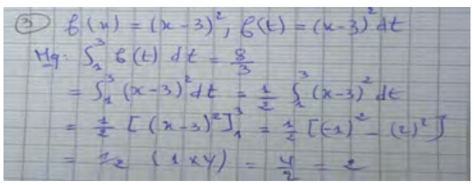
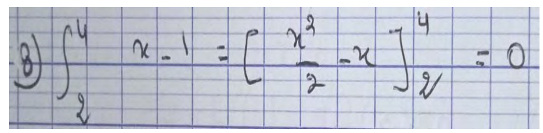
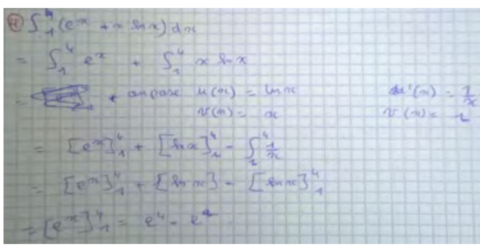
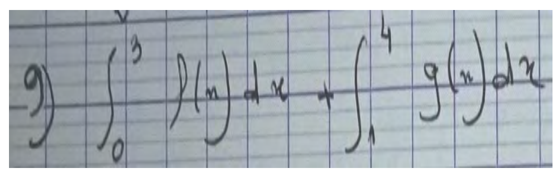
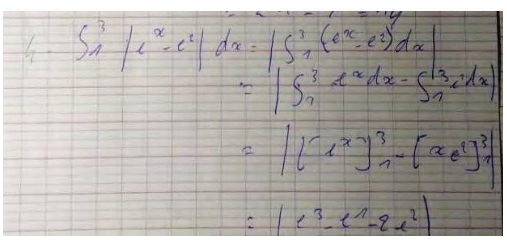
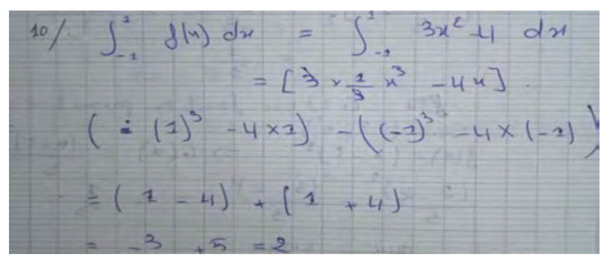
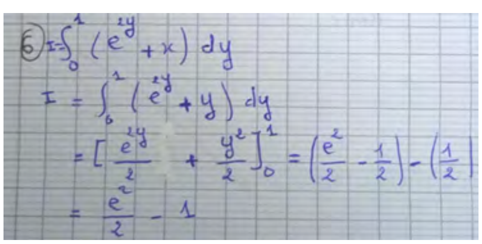
[24] Seah, E. K. (2005). Analysis of students' difficulties in solving integration problems. *The Mathematics Educator, 9(1)*, 39-59. http://math.nie.edu.sg/ame/matheduc/journal/v9_1/v91_39.aspx

[25] Sijmkens, E., Scheerlinck, N., De Cock, M., & Deprez, J. (2022). Benefits of using context while teaching differential equations. *International Journal of Mathematical Education in Science and Technology*, pp. 1-21. https://doi.org/10.1080/0020739X.2022.2039412

[26] Thomas, M., & Ye, Y. H. (1996). The Riemann integral in calculus: Students' processes and concepts. In P. C. Clarkson, *Proceedings of the 19th Annual Conference of the Mathematics Education Research Group of Australasia.* Australia. https://merga.net.au/Public/Publications/Annual_Conference_Proceedings/1996_MERGA_CP.aspx

[27] Watson, I. (1980). Investigating errors of beginning mathematicians. *Educational Studies in Mathematics, 11*, 319-329. https://doi.org/10.1007/BF00697743

APPENDIX: Examples of Students' Erroneous Productions

Question	Example of students' productions	Question	Example of students' productions
Q1		Q7	
Q3		Q8	
Q4		Q9	
Q5		Q10	
Q6		Q11