

## Analysis of a criterial examination of integral calculus: the case of a public university of Mexico

### Análisis de un examen criterial de cálculo integral: el caso de una universidad pública de México

DE LAS FUENTES-LARA, Maximiliano†, AGUILAR-SALINAS, Wendolyn Elizabeth\* and JUSTO-LÓPEZ, Araceli Celina

*Universidad Autónoma de Baja California, Faculty of Engineering, Mexico.*

ID 1<sup>st</sup> Author: Maximiliano, De Las Fuentes-Lara / ORC ID: 0000-0002-1001-4663 / CVU CONACYT ID: 412049

ID 1<sup>st</sup> Co-author: Wendolyn Elizabeth, Aguilar-Salinas / ORC ID: 0000-0003-2223-9234 / CVU CONACYT ID: 265060

ID 2<sup>nd</sup> Co-author: Araceli Celina, Justo-López / ORC ID: 0000-0002-6911-2063 / CVU CONACYT ID: 326795

DOI: 10.35429/EJS.2021.15.8.1.9

Received: July 10, 2021; Accepted: December 30, 2021

#### Abstract

This research analyzes the quality and results of a criterial and large-scale comprehensive calculus test in the school cycles between 2014 and 2019 to a total of 5367 second-semester students of the engineering careers of a Mexican public university. With the results obtained it is observed that the criterial examination of integral calculus is a valid, reliable test with satisfactory power of discrimination and with a majority load of procedural reagents. The results of the research show that the greatest difficulty for students is focused on integration techniques, especially when trigonometric functions are involved. It was also found that the success of students in the ECCI is due to the ability to resolve integrals with hyperbolic, exponential and logarithmic functions, as well as the proper application of the fundamental theorem of calculus and the technique of variable change.

#### Resumen

Esta investigación analiza la calidad y los resultados de un examen criterial y de gran escala de cálculo integral en los ciclos escolares comprendidos entre 2014 y 2019 a un total de 5367 estudiantes de segundo semestre de las carreras de ingeniería de una universidad pública mexicana. Con los resultados obtenidos se observa que el examen criterial de cálculo integral es una prueba válida, confiable y con poder de discriminación satisfactorio y con carga mayoritaria de reactivos procedimentales. Los resultados de la investigación evidencian que la mayor dificultad para los estudiantes está centrada en las técnicas de integración sobre todo cuando se involucran funciones trigonométricas. También se encontró que el éxito de los estudiantes en el ECCI se debe a la habilidad para la resolución de integrales con funciones hiperbólicas, exponenciales y logarítmicas, así como la aplicación adecuada del teorema fundamental del cálculo y la técnica de cambio de variable.

**Integral calculus, Criterial test, Evaluation**

**Cálculo integral, Examen criterial, Evaluación**

**Citation:** DE LAS FUENTES-LARA, Maximiliano, AGUILAR-SALINAS, Wendolyn Elizabeth and JUSTO-LÓPEZ, Araceli Celina. Analysis of a criterial examination of integral calculus: the case of a public university of Mexico. ECORFAN Journal-Spain. 2021. 8-15:1-9.

\* Correspondence of the Author (E-mail: aguilar.wendolyn@uabc.edu.mx)

† Researcher contributing as first author.

## Introduction

Subjects in the area of mathematics are an essential part of the professional training of the engineer. In the case of the Integral Calculus (CI) learning unit, it is intended that students appropriate the concept of integral and the techniques for its resolution. The study and learning of IC allow modeling, understanding and advancing in the knowledge of various phenomena of nature, such as: the area between curves, solids of revolution (volumes), arc length, centroids and problems of fall free; are common examples of it. For Tasman and Ahmad (2018), calculus is the heart of the curriculum and one of the most important learning units for science and engineering careers. Its objective is to provide students with knowledge that allows them to analyze, interpret, pose and solve science and engineering problems (Zúñiga, 2007), since the training of these professionals demands mathematical learning that contributes to solving specific problems of a technological nature, but above all practical (Ruiz, Carmona and Montiel, 2016).

Calculus is one of the most difficult but fundamental subjects in the training of engineering students (Machromah, Purnomo & Sari, 2019). For this reason, its theoretical study and understanding by future engineers is of the utmost importance.

It is known that in education, evaluation plays a very important role in determining school performance and the knowledge that students achieve, the consequence of which may be accountability, or the redirection of instructional design and teaching strategies.

At present, large-scale evaluation that is carried out through standardized instruments of national and international application is one of the most used tools to determine the knowledge and monitoring of student learning and the status of educational systems (Jiménez, 2016) With this type of evaluation, it is feasible to identify the topics or topics where students find greater complexity, to later design and implement strategies that allow to remedy or improve the quality of student learning.

Today there are large-scale assessment systems in almost all European Union countries and in other highly developed countries such as Japan, South Korea, Singapore and Israel. They are beginning to be implemented in Arab countries, and in Africa, with the support of UNESCO's International Institute for Educational Planning.

A positive consequence of the Program for International Student Assessment (PISA), and large-scale testing in general, is that it has drawn the attention of society to the importance of making efforts to improve the levels of learning that are achieved. reach in the educational system (Martínez-Rizo, 2016).

In the United States, for example, since the mid-20th century, large-scale tests have been developed to assess learning achievements in both the conceptual and procedural components, and their application is highly recommended to improve education systems (Bautista, 2015).

In Mexico, it is known that the results obtained in mathematics courses at practically any school level are not satisfactory, for Mexican students the results obtained in international assessments (PISA, 2012, 2015) are below the average Although there are many factors that intervene in such performances, deficiencies have been detected to represent and carry out treatments and conversions between the different registers of representation of mathematical concepts, which seems to be due to the preponderance of the teaching of mathematics based on a traditional approach, in which procedural content is emphasized. However, nowadays the information regarding the school effectiveness variables obtained is poor (López and Gamazo, 2020).

## Objective

A research study is carried out in a Faculty of Engineering of a public university in Mexico with the following particular objectives: (1) to present the criterial examination of integral calculation; (2) demonstrate that the ECCI is a valid, reliable assessment instrument with acceptable discriminatory power; (3) run a cluster analysis to identify the items and topics that predict student success; (4) carry out an analysis of the results obtained by the students in the ECCI.

## Theoretical framework

For Contreras and Backhoff (2004) large-scale exams are applied to hundreds or thousands of students and have the purpose of certifying educational achievement, they have direct consequences on the accreditation of students with respect to a course or subject.

The exam is declared criterial from the definition of the evaluative domain, when it comes to absolutely measuring the abilities of the students, what the examinee can do or not, serves as a diagnosis and significantly contributes to the creation or improvement of the design instructional, in the same vein, criterial tests are aimed at contrasting the curriculum designed, the one implemented and educational achievements (Jornet and González, 2009). Large-scale exams are usually designed with multiple choice (one correct answer and several distractors) in which the student is asked to choose the correct answer.

In the educational system, it is intended that students who attend schools are destined to acquire mainly theoretical and practical knowledge, so it is convenient to analyze whether the knowledge and content whose domain is evaluated refer to concepts or procedures.

For Sfard (1991), the structural or conceptual vision is characterized by an entity that has its own characteristics, is static, instantaneous and integrative. For Contreras and Backhoff (2004) it refers to the characteristics, ontological attributes and definition of the concept itself, it is also distinguished by the appropriation of a regularity. Harel, Selden and Selden (2006) consider it abstract and that it implies an implicit belief about the nature of mathematical entities that are described by formal definitions.

Regarding the operational or procedural vision, Sfard (1991) mentions that it is qualified by actions and algorithms, that it is dynamic and procedural or sequential. Contreras and Backhoff (2004) refer to them as the domain of the phases that implies the sequence of operations involved, the use of algorithms, and how, when and where they are used. since the learner is able to interpret multiple realities, he is better prepared to face real life situations. If a student can solve problems, he will be better prepared to apply her knowledge to new and changing situations (Guàrdia and Sangrà, 2005).

## Methodology

The data used in this research correspond to the results of the criterial integral calculus exam (ECCI) applied between the 2014 and 2019 school cycles. The ECCI was applied to a total of 5,367 second-semester students of engineering careers.

The administration of this exam, which consists of 40 multiple-choice items, was carried out in the computer labs of the same university and lasted a maximum of 2 hours. The answers are recorded and processed in the QuestionMark data platform.

The content validity is guaranteed from the selection of appropriate indicators related to mathematical processes and the contrast of the validity of the reagents through the judgment of experts (Alsina and Coronata, 2014).

A reliability analysis was performed using the Kuder-Richarson coefficient (KR-20), which allows obtaining the reliability of an instrument from the data obtained in a single application (Corral, 2009). Additionally, the Ferguson delta coefficient was calculated, which measures the discrimination power of a complete test, the range of this coefficient is [0.1] and it is satisfactory when it is greater than 0.90 (Ding et al., 2006).

The difficulty index (ID) was determined, which is related to the proportion of students who correctly solve a question, and is calculated according to Crocker and Algina, (1986). There are parameters for the acceptance of a reagent according to its level of difficulty, the one established by Contreras and Backhoff (2004), says that it must be greater than 0.05 and less than 0.95.

For Engelhardt (2009) the average of this index in the test must be 0.5 to maximize discrimination. According to Backhoff, Larrazolo and Rosas (2000), the average level of difficulty of the instrument should range between 0.5 and 0.6.

The discrimination index (IDC) of the item allows differentiating (discriminating) between those students who obtained good grades in the test and those who obtained low scores, it is related to the high possibility of correctly answering the item, that is, of those students with a overall outstanding performance in the test, the opposite situation in the case of students with poor performance, in this analysis 54% of the sample population is considered, since 27% of students with high performance and equal Percentage of students with the lowest performance, for each item that is reviewed.

For Contreras and Backhoff (2004) the discriminative value of the reagent is considered appropriate if it is greater than 0.2. The IDC scale according to Backhoff, Larrazolo and Rosas (2000) is: bad ( $IDC < 0.20$ ), fair ( $0.20 < IDC < 0.30$ ), good ( $0.30 < IDC < 0.40$ ) and excellent ( $IDC > 0.40$ ).

The correlation coefficient of the biserial point (rpbis) was also calculated. For some researchers (Henrysson, 1971; Molina, Wizner, Lacave and Gallardo, 2015) this coefficient is an indicator of predictive validity, where the response to an item by a student and the result obtained from the test is related. This psychometric indicator is calculated according to the Backhoff, Larrazolo and Rosas (2000) model. For Ding et al. (2006) the rpbis should be greater than 0.2 although only a few test items could not meet this condition.

The QuestionMark platform database contains the response to each item for each student, the item results are assessed as a correct (1) or incorrect (0) response, these results were analyzed through the IBM SPSS Statistics program and Microsoft® Office Excel. The information treatment allows obtaining the indices of difficulty, discrimination and the correlation coefficient of the biserial point, in addition to the reliability of the test by the Kuder Richardson (KR-20) and Ferguson delta methods.

An analysis of variance (ANOVA) was carried out with a post-hoc Tukey HSD test, in which the number of item was considered as a factor and the ID as a dependent variable in order to determine significant differences between the five units that make up the integral calculus course.

An analysis was also carried out on the development of reagent profiles, in order to identify the reagents and topics that predict student success and the existing relationship with discrimination and difficulty, for this purpose, the technique called cluster analysis (Bausela, 2005; Castejón et al., 2016, Dixson et al., 2017; Gonçalves et al., 2017). This analysis is a type of data classification that is carried out by grouping the analyzed elements. The fundamental objective of this type of analysis is to classify  $n$  objects into  $k$  ( $k > 1$ ) groups, called clusters, by using  $p$  ( $p > 0$ ) variables. The type of classification is  $k$ -means, as it is a tool designed to assign cases to a fixed number of groups.

## Results and discussion

This section has been divided into four parts, the first refers to the presentation of the ECCI, followed by the quality analysis of the ECCI, the cluster analysis and finally the analysis of the results obtained by the students in the ECCI.

### *ECCI presentation*

The ECCI is formally applied in the Faculty of Engineering of a public university in Mexico to all students who take the Comprehensive Calculus learning unit. The results of this exam are part of your ordinary evaluation and provide 30% of the final grade for this course, while the other 70% is awarded by the teacher who teaches the subject. The ECCI has been applied since 2005, it is currently made up of 40 items and has the following characteristics: it is criterial, it is aligned with the curriculum, it is multiple choice (one correct answer and three distractors) and it is large-scale.

In the construction of the ECCI, the model of Nitko (1994) was adopted to develop exams oriented by the curriculum, this model was complemented by the methodology for the construction of criterial tests of Popham (1990) and with methodological and operational contributions of Contreras (1998, 2000). The ECCI explores five basic units of knowledge of the Integral Calculus course, the items that make up each unit and their specific competence are shown in table 1.

Unit	Name	Competence	Reactives
1	Anti-derivation and definite integral	Calculate the antiderivative of a function and its definite integral by definition, using the corresponding theorems, to discern about the use and application of the concept of integral, with a critical, proactive and responsible attitude	1-7
2	Applications of the integral	Solve geometric engineering problems, based on the use of mathematical theorems and models, to design, optimize engineering processes and systems, with a critical and responsible attitude.	8-11

3	Transcendent functions	Calculate integrals of transcendent functions, for the resolution of problems that involve the analytical, graphical and numerical aspects, through their properties and theorems, with a disposition for teamwork, a critical and responsible attitude	12-20
4	Integration techniques	Solve definite and indefinite integrals, by identifying and using the corresponding integration techniques, for application in various engineering problems, with a disposition for collaborative work, a critical and responsible attitude.	21-32
5	Improper integrals	Solve geometric problems with improper integrals, applying the concept of limit, to design, optimize engineering processes and systems, with a critical, proactive attitude and a willingness to collaborate work	33-40

**Table 1** Competence and reagents of the differential calculus exam

Source: *Self made*

For the design of each of the reagents, a technical specification was previously developed based on the content of integral calculation and the psychometric point of view. The specification consists of a section of data identifying the content to be evaluated, an interpretation of its meaning, the characteristics of the stimuli that the item will present to the examinee (instructions to respond, item base, distractors, correct response, tables or figures of support, etc.), as well as a sample reagent that exemplifies the application of such reagent design considerations.

### *ECCI quality analysis*

The existing documentation (content and grid of the subject, justification of the contents, table of specifications and specifications of the reagents) on the design and construction of the ECCI evidences the selection of adequate indicators related to the mathematical processes and the contrast of the validity of the reagents through the judgment of experts (Alsina and Coronata, 2014). The content validity was guaranteed with the participation of expert judges (Integral Calculus teachers with a minimum teaching experience of 5 years) and the group consensus method (Corral, 2009). In the subjects object of the validation, they were those who reviewed the coherence of the items with what is to be evaluated, the complexity of the items and the cognitive ability to be evaluated (Barraza, 2007), as well as the sufficiency and relevance of the items, in where the relevant aspects of the construct were considered, included in the competencies and indicators (Cisneros, Jorquera and Aguilar, 2012).

The average reliability during the 6 years of application was calculated using the Kuder Richardson method (KR-20), a coefficient  $r = 0.86$  was obtained, which is considered appropriate when it is greater than or equal to 0.85 in the case of standardized criterion tests. and large-scale (Contreras and Backhoff, 2004; Ding et al., 2006; Muñoz and Mato, 2008). The average distribution of the total scores was also calculated using the Ferguson delta test, obtaining a value of 0.98, which broadly satisfies the established criteria (Ding, et al., 2006; Engelhardt, 2009).

The average value of the difficulty index is  $0.51 \pm 0.11$  (mean  $\pm$  standard deviation), it is a medium and adequate level of difficulty (Backhoff, Larrazolo and Rosas, 2000; Ding et al., 2006; Engelhardt, 2009). No reagent with ID less than 0.2 or greater than 0.85 was identified. In contrast, the minimum average value of the ID is 0.35 while the maximum average value of the ID is 0.77, the difficulty is appropriate according to the criteria of the specialists.

The average of the IDC is  $0.44 \pm 0.09$  (mean  $\pm$  standard deviation), which falls within a rating considered excellent (Backhoff, Larrazolo and Rosas, 2000), it also meets the criteria declared by Ding et al. (2006) in the which the average IDC is greater than 0.3.

The average of the biserial correlation coefficients of the test is  $0.35 \pm 0.09$  (mean  $\pm$  standard deviation), the coefficient complies with the recommendation of the specialists (Ding et al., 2006; Engelhardt, 2009) in that the mean of the rpbis is greater than 0.2.

### Cluster analysis

To establish the significant characteristics between the psychometric indicators and the results of the students, a cluster analysis of k-means was carried out. The results are three profiles that are described in table 2.

Psychometric indicators	Group		
	1	2	3
Difficulty index	0.58	0.60	0.41
Discrimination index	0.51	0.38	0.34
rpbis	0.44	0.33	0.29
Amount of reagents	13	9	18

**Table 2** Final cluster centers

Source: Self made

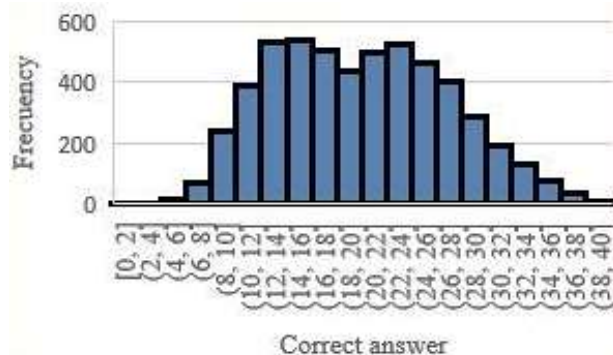
There are 13 items in group 1 and are characterized by their medium difficulty (0.58), by the excellent discrimination value (0.51) and items with the highest prediction (0.44) compared to the rest of the groups. They are reagents that mostly belong to units 1 and 3, the method of change of variable, the application of the fundamental theorem of calculus, the resolution of integrals with hyperbolic, exponential, logarithmic and trigonometric functions stand out as topics in this group.

The items in group 2 are 9 and are characterized by their medium difficulty (0.60), the discrimination value (0.38) and items with acceptable rpbis (0.33). It is observed that the reactants that belong to this group correspond to the calculation of area under the curve, area in two regions, solids of revolution and integration techniques.

Finally, group 3 consists of 18 items, with the most difficult items (0.41), acceptable discrimination (0.34) and the lowest biserial correlation coefficient (0.29). In this group there are reagents from all units, but unit 4 (integration techniques) stands out, since 44% of the reagents belong to that unit.

### Analysis of the results obtained by the students in the ECCI

The average of the correct answers in the ECCI is 20.48 out of 40 possible points (graph 1), the average expressed as a percentage of the total points is 51%, which corresponds to the average ID of 0.51.



**Graphic 1** Distribution of the number of correct answers in the ECCI

Source: Self made

The distribution of the number of correct items was significantly non-normal (Kolmogorov-Smirnov,  $D(5367) = 0.076$ ,  $p < 0.000$ ). The asymmetry of the distribution of the number of correct items is 0.218 (Deviation = 0.033), these values indicate an asymmetry positive, the kurtosis of the distribution is -0.678 (Deviation = 0.067). For the type of distribution, the measures of central tendency and dispersion are added, the median is 20, the mode is 15, the quartile 1 is 15, the quartile 2 is 20 and quartile 3 is 26 (25% obtained a score equal to or greater than 65). Students had difficulty with 21 items (rated as highly and moderately difficult) and the positive asymmetry shows the difficulty of the ECCI, only the 34.6% obtained a score equal to or greater than 60 (on a scale of 0 to 100) with which they accredit the ECCI, that is, of the 5,367 students who applied the ECCI only 1,859 approved it.

Table 3 describes the characteristics of the 5 most difficult items for students, according to ID. These reagents belong to the techniques of integration by parts, trigonometric functions and case 4 partial fractions called repeated quadratics, all the reagents have content of the procedural type.

Reactive	Contents	Achievement indicator	ID
22	Procedural	Solve indefinite integrals by using the so-called piecemeal technique	0.35
24	Procedural	Solving definite and indefinite integrals of powers of secant and tangent	0.36
25	Procedural	Solve definite and indefinite integrals by using the cosecant and cotangent powers technique	0.37
23	Procedural	Solve definite and indefinite integrals of powers of sine and cosine	0.38
32	Procedural	Solve definite and indefinite integrals by using the technique called	0.38

**Table 3** The five most difficult questions

Source: Self made

Table 4 describes the characteristics of the 5 items with the highest prediction, according to the biserial correlation coefficient. It was found that the most predictive items of the success of the students in the ECCI are of the procedural type, involving the resolution of integrals with hyperbolic, exponential and logarithmic functions, the application of the fundamental theorem of calculus and the variable change technique that is practically use throughout the course.

Reactive	Contents	Achievement indicator	rpbis
19	Procedural	Solving Indefinite Integrals of Hyperbolic Functions	0.50
12	Procedural	Solve indefinite integrals of exponential and / or logarithmic functions.	0.48
7	Procedural	Apply the Fundamental Theorem of Calculus to Solve Definite Integrals	0.46
4	Procedural	Solving integrals using the change of variable method	0.46
16	Procedural	Solve indefinite integrals that result in sine, tangent, or inverse secant.	0.45

**Table 4** The five most predictive items

Source: Self made

The analysis of variance between the items of each of the units and the ID did not show significant differences ( $p = 0.131$ ). The averages of the ID for the five units calculated are: 0.56 for unit 1 (antiderivation and definite integral), for unit 2 (applications of the integral) it is 0.55, unit 3 (transcendent functions) has an average ID of 0.56, in unit 4 (integration techniques) the average difficulty is 0.45 and 0.49 for unit 5 (improper integrals). With these values, congruence is observed with the reactants with greater difficulty that belong to integration techniques and involve trigonometric functions and rational functions.

In the ECCI, it was found that only 5 items of the 40 (12.5%) are conceptual, while 35 (87.5%) are procedural, the above is a reliable example of the emphasis of a teaching scheme that does not show a balance between the conceptual and procedural contents. A T-test of independent samples between the conceptual cut reagents (ID = 0.50) and the procedural reagents (ID = 0.51) did not show significant differences regarding the difficulty of the ECCI reagents.

## Conclusions

The quality of the ECCI that is applied on a large scale in a public university in Mexico was reviewed and whose main characteristics are: criterial, aligned to the curriculum, multiple choice and high impact, its application began in 2005 and consists of 40 items of the of which 87.5% are procedural.

It was determined that the ECCI has valid content, is reliable and has satisfactory discrimination power.

Reliability was calculated using the KR-20 method and a coefficient  $r = 0.86$  was obtained, the Ferguson delta coefficient was also determined, whose resulting value is 0.98, which are considered by specialists as satisfactory.

The asymmetry of the distribution of the number of correct items evidenced the difficulty of the ECCI with an average ID = 0.51, which is practically at the lower limit of the range established by Backhoff, Larrazolo and Rosas (2000).

The ECCI revealed that the greatest difficulty for students is centered on integration techniques, especially when trigonometric functions are involved.

The conglomerate analysis carried out allowed to establish that the predictive items of a student's success in the ECCI are associated with the understanding and integration of hyperbolic, exponential and logarithmic functions, the correct use of the fundamental theorem of calculus and the adequate use of the change of variable, also called substitution technique.

The percentage of procedural reagents is latent evidence of the teaching scheme that prevails in the Faculty of Engineering of this public university, the incorporation of technological resources, the use and balance of different semiotic representations, as well as the search for a balance between procedural content. The revision and updating of the ECCI is an imperative need from the perspective of the results found. Today students need to explore, visualize, conjecture, analyze and interpret rather than carry out a series of step-by-step algorithms to obtain a result.

## References

- Alsina, Á. y Coronata, C. (2014). Los procesos matemáticos en las prácticas docentes: diseño, construcción y validación de un instrumento de evaluación. *Educación Matemática En La Infancia*, 3(2), 23–36.
- Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del examen de habilidades y conocimientos básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1), 1-19.
- Barrazas, A. (2007). La consulta a expertos como estrategia para la recolección de evidencias de validez basadas en contenido. *Investigación Educativa Duranguense*, 7, 5-13.
- Bausela, E. (2005). SPSS: un instrumento de análisis de datos cuantitativos, *Revista de Informática Educativa y Medios Audiovisuales*, 2(4), 62-69.
- Muñoz, J. M. y Mato, M. D. (2008). Análisis de las actitudes respecto a las matemáticas en alumnos de ESO. *Revista de Investigación Educativa*, 26(1), 209-226.
- Castejón, J. L., Gilar, R., Miñano, P., y González, M. (2016). Latent class cluster analysis in exploring different profiles of gifted and talented students. *Learning and Individual Differences*, 50, 166-174. doi: <http://doi.org/10.1016/j.lindif.2016.08.003>
- Cisneros, E., Jorquera, M. y Aguilar, A. (2012). Validación de instrumentos de evaluación docente en el contexto de una universidad española. *Voces y Silencios: Revista Latinoamericana de Educación*, 3(1), 41-55.
- Contreras, L. A. (1998). *Metodología para desarrollar y validar un examen de español, de referencia criterial y referencia normativa orientado por el currículum, para la educación primaria en México*, en Tercer Foro Nacional de Evaluación Educativa. Memorias. Veracruz: CENEVAL.
- Contreras, L. A. (2000). *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California*. Tesis. Recuperado de: [http://ide.ens.uabc.mx/documentos/divulgacion/tesis/MCE/1998/Luis\\_Angel\\_Contreras\\_Nino.pdf](http://ide.ens.uabc.mx/documentos/divulgacion/tesis/MCE/1998/Luis_Angel_Contreras_Nino.pdf)
- Contreras, L. y Backhoff, E. (2004). *Educación, aprendizaje y cognición. Teoría en la práctica*. México: Manual moderno.
- Corral, Y. (2009). Validez y confiabilidad de los instrumentos para la recolección de datos. *Revista ciencias de la educación*, 33, 228-247.
- Crocker, L. y Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Florida: Holt, Rinehart and Winston.
- Ding, L., Chabay, R., Sherwood, B. y Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical review special Topics-Physics education research*, 2(1), 1-7.
- Dixson, D. D., Worrell, F. C. y Mello, Z. (2017). Profiles of hope: How clusters of hope relate to school variables. *Learning and Individual Differences*, 59, 55-64. doi: <http://doi.org/10.1016/j.lindif.2017.08.011>
- Goncalves, T., Niemivirta, M. y Lemos, M. S. (2017). Identification of students' multiple achievement and social goal profiles and analysis of their stability and adaptability. *Learning and Individual Differences*, 54, 149-159. doi: <http://doi.org/10.1016/j.lindif.2017.01.019>.
- Harel, G., Selden, A. y Selden, J. (2006). Advanced mathematical thinking: Some PME perspectives. In *Handbook of research on the psychology of mathematics education*, 147-172. Brill Sense.
- Engelhardt, P. V. (2009). An introduction to classical test theory as applied to conceptual multiple-choice tests. *Getting started in PER*, 2(1).
- Guàrdia, L. y Sangrà, A. (2005). Diseño instruccional y objetos de aprendizaje; hacia un modelo para el diseño de actividades de evaluación del aprendizaje on-line. *RED. Revista de Educación a Distancia*, 4, 1-14. Recovered from: <https://www.um.es/ead/red/M4/guardia17.pdf>
- Henrysson, S. (1971), *Gathering, Analysing, and Using Data on Teste Items*, en R.L. Thorndike (Ed.), *Educational Measurement*. Washington, DC: American Council on Education.
- Jiménez, J. A. (2016). El papel de la evaluación a gran escala como política de rendición de cuentas en el sistema educativo mexicano. *RIEE. Revista Iberoamericana de Evaluación Educativa*, 9(1), 109-126. doi: <https://doi.org/10.15366/riee2016.9.1.007>



- López, A. y Gamazo, A. (2020). Estudio multinivel sobre las variables explicativas de los resultados de México en PISA 2015. *Education Policy Analysis Archives*, 28(26), 1-22.
- Machromah, I. U., Purnomo, M. E. R., & Sari, C. K. (2019, February). Learning calculus with geogebra at college. In *Journal of Physics: Conference Series 1180*, 1-11. doi: <https://doi.org/10.1088/1742-6596/1180/1/012008>
- Martínez-Rizo, F. (2016). Impacto de las pruebas en gran escala en contextos de débil tradición técnica: Experiencia de México y el Grupo Iberoamericano de PISA. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 22(1), 1-12.
- Jornet, J. M. y González, J. (2009). Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. *Estudios sobre Educación*, 16, 103-123.
- Molina, A. I., Wizner, Á. L., Lacave, C. y Gallardo, J. (2015). Una herramienta de diseño y análisis de instrumentos de evaluación e indagación docente. *Actas de las XXI Jornadas de la Enseñanza Universitaria de la Informática*, 144-151. Universitat Oberta La Salle.
- Nitko, A. J. (1994). *A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students*. Conferencia Internacional sobre Evaluación y Medición Educativas, de la Asociación para el Estudio de la Evaluación Educativa (ASSESA).
- PISA. (2012). *Resultados de PISA 2012*, Grupo Fórmula. Recovered from: <https://www.oecd.org/pisa/keyfindings/PISA-2012-results-mexico-ESP.pdf>.
- PISA. (2015). *Resultados de PISA 2015*, Grupo Fórmula. Recovered from: <https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf>.
- Popham, J. (1990). *Modern educational Measurement: a practitioner's perspective*. Boston: Prentice Hall.
- Ruiz, E. F., Carmona, E. A. y Montiel, A. S. (2016). Importancia del cálculo en el desarrollo académico del ingeniero. *Pistas Educativas*, 120, 402-420.
- Bautista, E. J. (2015). La evaluación mediante pruebas de gran escala en México. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 5(10), 1-15.
- Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22, 1-36.
- Tasman, F. y Ahmad, D. (2018). Visualizing Volume to Help Students Understand the Disk Method on Calculus Integral Course. In *IOP Conference Series: Materials Science and Engineering*, 335(1), 1-6. IOP Publishing.
- Zúñiga, L. (2007). El cálculo en carreras de ingeniería: un estudio cognitivo. *Revista Latinoamericana de Investigación en Matemática Educativa*, 10(1), 145-175.