A measurement instrument for establishing the algebraic skills of engineering students on a Differential Calculus Course in engineering

Instrumento de medición para diagnosticar las habilidades algebraicas de los estudiantes en el Curso de Cálculo Diferencial en ingeniería

Wendolyn Elizabeth AGUILAR-SALINAS, PhD. Associate Professor. Universidad Autónoma de Baja California *(aguilar.wendolyn@uabc.edu.mx).*

Maximiliano DE LAS FUENTES-LARA, PhD. Associate Professor. Universidad Autónoma de Baja California *(maximilianofuentes@uabc.edu.mx).*

Araceli Celina JUSTO-LÓPEZ, PhD. Associate Professor. Universidad Autónoma de Baja California *(araceli.justo@uabc.edu.mx).*

Ana Dolores MARTÍNEZ-MOLINA, M.E. Associate Professor. Universidad Autónoma de Baja California *(ana.dolores.martinez.molina@uabc.edu.mx).*

Abstract:

A highly reliable algebraic skill measurement instrument with content and approach validity was developed. Its content focusses on the algebraic skills engineering students require to successfully follow a Differential Calculus Course. A team of 10 teachers, each with minimum of a master's degree and teaching experience in differential calculus, participated in the design of this instrument. The measurement instrument is a large-scale multiple-choice criteria test comprising 25 test items. Its quality is described and analysed on the basis of the answers given by engineering

Revision accepted: 2019-09-18.

This is the English version of an article originally printed in Spanish in issue 275 of the **revista española de pedagogía**. For this reason, the abbreviation EV has been added to the page numbers. Please, cite this article as follows: Aguilar-Salinas, W. E., de las Fuentes-Lara, M., Justo-López, A. C., & Martínez-Molina, A. D. (2020). Instrumento de medición para diagnosticar las habilidades algebraicas de los estudiantes en el Curso de Cálculo Diferencial en ingeniería | *A measurement instrument for establishing the algebraic skills of engineering students on a Differential Calculus Course in engineering. Revista Española de Pedagogía*, *78* (275), 5-25. doi: https://doi.org/10.22550/REP78-1-2020-02

students during the first and second semesters of the 2018-2019 academic year. The results show that topics that can predict student success and have the greatest power of discrimination in the measurement instrument are strongly related to skills students acquire in primary and secondary education, such as operating with fractions and the laws of exponents. It was also found that the main shortcomings in the algebraic skills of students are rationalisation, division of polynomials, factoring sums, and difference of cubes.

Keywords: calculus, questionnaire, predictive evaluation, reliability and validity.

Resumen:

Se construyó un instrumento de medición altamente confiable, con validez de contenido y de criterio. Su contenido está basado en las habilidades algebraicas que los estudiantes de ingeniería requieren para desempeñarse favorablemente en un Curso de Cálculo Diferencial en las carreras de ingeniería. En el diseño del instrumento participó un equipo de 10 profesores con al menos grado de maestría y experiencia docente en el área de cálculo diferencial. El instrumento de medición es de opción múltiple, criterial, de gran escala, está integrado por 25 reactivos y su análisis de calidad se describe y se deriva de las respuestas emitidas durante los ciclos lectivos 2018-2 y 2019-1 por estudiantes de nuevo ingreso en la carrera de ingeniería. Los resultados muestran que los tópicos que predicen el éxito del alumno y cuentan con el mayor poder de discriminación en el instrumento de medición están fuertemente relacionados con habilidades que los estudiantes adquieren desde la primaria y secundaria, como es el caso de las operaciones con fracciones y las leyes de los exponentes. También se logró identificar que la mayor deficiencia en las habilidades algebraicas de los estudiantes pertenece al tema de la racionalización, división de polinomios, factorización de suma y diferencia de cubos.

Descriptores: cálculo, cuestionario, evaluación predictiva, fiabilidad, validez.

1. Introduction

A command of mathematics is a vital skill in a society undergoing unprecedented technological development. However, for many students it is one of the most inaccessible skills as it comprises a large number of difficulties and failures (Carbonero & Navarro, 2006), making mathematics a critical filter that shapes students' choice of degree (Sells, 1973). Students who start university with a negative perception of mathematics develop a reluctant attitude and consequently have low academic performance, and attribute their failure to a variety of factors (Orozco & Díaz, 2009).

In Mexico, mathematics levels are considered to be an educational problem, with 50% of students in early stages of their education showing a lack of interest in this area (González, 2005). This is reflected in what Mexico's Department of Public Education (Secretaría de Educación Pública, SEP) has published in coordination with the National Institute for Evaluating Education (Instituto Nacional para la Evaluación de la Educación, INEE) and the educational authorities of Mexico's different states, in which they make use of the test of the National Plan for Evaluating Education (Plan Nacional para la Evaluación de los Aprendizajes, PLANEA) for secondary education to reflect the low performance of baccalaureate students in the area of mathematics at a national level in the three years between 2015 and 2017 (Table 1). This shows that the largest percentage of Mexican students are at level one, where they perform operations with fractions and operations that combine unknowns and establish and analyse relationships between two variables.

TABLE 1. Figures in percentages from the PLANEA test.

Year/Level	2015	2016	2017
Level 1	51.3	49.2	66.2
Level 2	29.9	30.0	23.3
Level 3	12.4	14.4	8.0
Level 4	6.4	6.3	2.5

Source: Own elaboration.

It is necessary to establish the students' initial level of knowledge to enable them to learn and avoid making assumptions. For example, mathematics is a vital subject in the training of engineers (Morales, 2009); if they do not have basic knowledge of mathematical skills owing to poor training in the baccalaureate, then they will struggle to understand and assimilate university-level mathematics (Encinas, Osorio, Ansaldo, & Peralta, 2016). Evidence of deficient pre-university education is found in the students' poor performance on university courses (Orozco & Diaz, 2009).

As a result, studies have analysed personal, sociodemographic, psychological, intellectual and cognitive factors and even the academic records of the students (Reynoso & Méndez-Luévano, 2018; Arriaga, 2015; González, 2013; Gatica-Lara, Méndez-Ramírez, Sánchez-Mendiola, & Martínez-González, 2010; Difabio, 1994). However, we did not find research analysing the topics or items that could affect the students' success in the area of mathematics.

The students' abilities are not constructed alone, but rather on the basis of prior knowledge. Accordingly, Ausubel, Novak, and Hanesian (1983, p. 1) observe that «the most important factor influencing learning is what students already know. Find this out and teach them accordingly.» This is apparent in the study by Orozco-Moret and Morales (2007), where 70% of the students who were repeating a mathematics module in the first semester of university agreed that lack of previous knowledge was the main cause of their problem with the module.

In view of the above, this research refers to the construction and validation of an instrument to measure the algebra skills university students require to perform adequately in a Differential Calculus Course on engineering degrees at the Universidad Autónoma de Baja California (UABC). Establishing what algebra skills they have acquired during their education at lower levels is crucial for the students' success on a differential calculus for engineering course.

2. Method

To construct this measurement instrument, we used the Nitko model (1994) for developing exams shaped by the curriculum. This model is complemented by Popham's (1990) methodology for constructing criterion-referenced tests and by methodological and operational contributions from Contreras (1998, 2000).

The quality of the measurement instrument was analysed in accordance with classic test theory (CTT) to ensure the instrument allows for measurement of the algebra skills required to take the Differential Calculus Course on an engineering degree. Accordingly, it is necessary to determine its reliability, validity, indexes of difficulty and discrimination, and the biserial correlation (Carmines & Zeller, 1987).

The analyses of reliability allow us to measure the consistency or stability of the measurements when the measurement process is repeated (Prieto & Delgado, 2010), determining its capacity to demonstrate stability in its results (García & Vilanova, 2008). To do this, we used the Kuder-Richardson KR-20 formula and the split-half method.

Reliability analysis using the Kuder-Richardson (KR-20) formula makes it possible to establish an instrument's reliability based on the data obtained in a single application. The items are evaluated as correct or incorrect answers and the fact the items have different indexes of difficulty is taken into consideration (Corral, 2009). In the split-half method reliability analysis, the test is split in half and separated into two different parallel tests. The internal consistency coefficient is calculated using the Spearman-Brown formula (Reidl-Martínez, 2013). If the instrument is reliable, there should be a strong correlation between the scores in both halves.

Content and criterion validity were also calculated in turn to analyse the quality of the instrument. The content validity is guaranteed by choosing indicators that are appropriate and related to the mathematical processes, as well as and by testing the validity of the items by expert judgment (Alsina & Coronata, 2014). In this type of validation, a panel of experts is selected. These experts have at least 5 years' experience in the topics being validated and will analyse whether the items are consistent with what they are intended to evaluate, the complexity of the items, and the cognitive skill being evaluated (Barrazas, 2007), as well as the sufficiency and relevance of the items, considering the aspects of the construct that are relevant, including those in the competences and indicators (Cisneros, Jorquera, and Aguilar, 2012).

The measurement instrument designed is intended to establish the students' knowledge or command of the algebra content or topics considered necessary for studying and handling differential calculus on engineering programmes. With the aim of determining whether the items in the measurement instrument actually examine the topics and indicators of achievement established in the design specifications, the measurement instrument was reviewed by a panel of six expert university teachers from the field of mathematics with master's or doctoral level degrees who were not connected with the process of designing and constructing the instrument. The expert panel evaluated the measurement instrument using a questionnaire which included 8 significant aspects for each of the 25 items. A scale of 0 to 4 was used to evaluate each aspect and item, with 0 indicating strongly disagree, 1 disagree, 2 neutral position, 3 agree, and 4 strongly agree.

Aspect 1 establishes the relevance of the content of the reactive to topics covered in level one of higher education. Aspect 2 measures whether the content of each item comprises topics covered in the introductory algebra, geometry, and trigonometry course for incoming students at the Mexicali Faculty of Engineering (FIM) at the UABC. This course is taken by incoming students who will study engineering programmes before the start of the academic year. The algebra topics covered are real numbers, exponents, radicals, fractions, rationalisation, polynomials, and factoring. Aspect 3 relates to the consistency between the indicator of achievement and the thematic content of the item. Aspect 4 considers the relevance of the content of the item to the requirements a student must fulfil to take the Differential Calculus Course in the FIM. Aspect 5 identifies whether the vocabulary used in each item is commonly used in the subject. Aspect 6 considers whether the distractors in each item are plausible. Aspect 7 considers whether each item has the correct answer. And aspect 8 refers to whether what is asked of the student is clear in each item.

Criterion validity refers to the extent to which the test correlates to variables external to it, which are called criteria. Consequently, the criterion is an indicator of what the test is intended to measure or of what should present a given relationship. The correlation found is called the coefficient of validity. The differential calculus baseline test was used to test criterion validity. The design and construction of this baseline test is described in Encinas, de las Fuentes, and Rivera, (2007) and in Contreras, Encinas, de las Fuentes, and Rivera (2005) and it is applied formally at the UABC to all students who take differential calculus module (more than 3000 students per semester distributed across the campuses in the state of Baja California, Mexico). The results of this exam form part of their ordinary evaluation, representing 30% of the final grade for the Differential Calculus Course with the other 70% being allocated by the teacher who delivers the module. The differential calculus baseline test, which has been in use since 2005, currently comprises 60 items and is aligned with the curriculum. It is criterion-referenced, multiple choice, and large scale.

The measurement instrument is classed as a criterion-referenced test as it is intended to establish algebra skills and support the diagnosis of the teaching design for the Differential Calculus Course. The difficulty index (DI) relates to the proportion of students who correctly solve an item. It is calculated in accordance with Crocker and Algina (1986). There are parameters for accepting an item depending on its level of difficulty. The difficulty index set by Contreras, Backhoff, and Larrazolo (2004) states that it must be greater than 0.05 and less than 0.95. For CTT this index must be between 0.1 and 0.9. According to Backhoff, Larrazolo, and Rosas (2000) the mean level of difficulty of the instrument should be between 0.5 and 0.6, with the values of the difficulty index being distributed as follows: 5% very easy items ($0.87 < DI < 1$), 20% fairly easy ($0.74 < DI < 0.86$), 50% with an average level of difficulty ($0.53 < DI < 0.73$), 20% fairly difficult ($0.33 < DI < 0.52$), and 5% very difficult ($DI < 0.32$).

The index of discrimination (IDC) of the item makes it possible to differentiate (discriminate) between students who obtained high scores in the test and those who obtain low scores. It relates to the high possibility that students with an excellent overall performance in the test will correctly answer the item, while students with poor performance will not. This analysis considers 54% of the sample population, including 27% of the students with high performance and the same percentage of students with the lowest performance for each item reviewed. Contreras, Backhoff, and Larrazolo (2004), and CTT regard an item's discrimination value as suitable if it is greater than 0.2. According to Guilford (1975), the index of discrimination of an item is accepted if it has a value greater than 0.2 or 0.3. The IDC scale according to Backhoff, Larrazolo, and Rosas (2000) is: poor ($IDC < 0.20$), mediocre ($0.20 < IDC < 0.30$), good ($0.30 < IDC < 0.40$), and excellent ($IDC > 0.40$).

Also regarded as important for the reliability and validity of the instrument is the point biserial correlation (rpbis), as this considers 100% of the sample population, not just 54% as in the case of the index of discrimination. According to Henrysson (1971), this coefficient is an indicator of predictive validity where a student's response to an item and result in the test are related. It is calculated in accordance with the model of Backhoff, Larrazolo, and Rosas (2000) and the scale of values for this indicator is: poor discrimination ($rpbis < 0.14$), mediocre discrimination ($0.15 < rpbis < 0.25$), good discrimination ($0.26 < rpbis < 0.35$), and excellent discrimination ($rpbis > 0.35$).

The results for each answer option are subjected to a frequency analysis in which the percentages of students who gave each of the four answer options is identified. Distractors are classed as unsuitable (UD) if they do not get over 5% of the answers from students (Rodríguez, Casas, & Medina, 2005). An analysis of variance (ANOVA) test was performed along with a *post-hoc* Tukey HSD test, in which the number of UDs was taken as the factor and the UDs, IDCs, and the rpbis of the items as the dependent variables. The objective of this was to identify significant differences between and within groups.

The database was analysed with classical test theory (CTT) using the IBM SPSS Statistics 25 program and Excel spreadsheet program with which we found the psychometric data for each item, distractors, difficulty indexes, discrimination indexes, and the point biserial correlation.

3. Process for constructing the measurement instrument

Six lecturers took part in the construction of the measurement instrument: two in the instrument design committee, two in the specifications preparation committee, and two in the item preparation committee. All of them held doctorates and had a minimum of 5 years' teaching experience in the area of algebra, differential calculus, and integral calculus.

The role of the measurement instrument design committee was to analyse the area's curriculum, detect and structure the important content that was to be evaluated, create a table of specifications for the instrument, and prepare a document to explain their decisions. It should be noted that the measurement instrument is based on the minimum arithmetic and algebraic skills engineering students require to successfully complete a Differential Calculus Course as part of their engineering studies. These skills were determined by the instrument design committee and validated by the specifications preparation committee and by the panel of experts. The mathematical concepts and procedures covered by the measurement instrument are part of the curriculum the students followed throughout their studies on the programme of the Mathematics I course on both the general baccalaureate and the technological baccalaureate (SEP, 2017). The topics are also covered in the algebra, geometry, and trigonometry course that students take when starting engineering degrees at the Universidad Autónoma de Baja California.

The aim of the measurement instrument is to establish the starting conditions of newly-enrolled students on engineering degrees regarding the algebra knowledge and skills required to complete the differential calculus module. To reflect these conditions, we established indicators of achievement for each specification and its respective item, representing the student behaviours that make it possible to evaluate the level of command of a particular algebra skill. The topics and indicators of achievement for each of the 25 items that comprise the measurement instrument are described in Table 2.

Table 2. Topics and indicators of achievement for each item in the measurement instrument.

Item number	Topic number	Topic	Indicators of achievement
1	1	Operations with fractions	Adding fractions with different denominators.
2	1	Operations with fractions	Multiplying fractions.
3	1	Operations with fractions	Dividing two fractions.
4	2	Laws of exponents	Using the laws of exponents to multiply two numbers.
5	2	Laws of exponents	Using the laws of exponents to divide two numbers.
6	2	Laws of exponents	Converting a radical number to an exponential.
7	3	Rationalisation	Rationalising the denominator of a numerical expression.
8	3	Rationalisation	Rationalising the denominator of an algebraic expression.
9	3	Rationalisation	Rationalising the numerator of a numerical expression.
10	4	Operations with polynomials (sum)	Calculating the sum of polynomials.
11	4	Operations with polynomials (sum)	Eliminating grouping symbols and simplifying the algebraic expression.
12	4	Operations with polynomials (product)	Calculating the product of two binomials.
13	4	Operations with polynomials (division)	Calculate the division of a polynomial by a monomial.
14	4	Polynomial operations (division)	Calculate the division of a polynomial by a binomial.
15	4	Operations with polynomials	Isolating a variable.
16	4	Operations with polynomials (product of binomials)	Squaring a binomial.

17	4	Operations with polynomials (binomial conjugates)	Calculating the product of binomial conjugates.
18	4	Polynomial operations (product of binomials)	Calculating the product of two binomials.
19	4	Operations with polynomials (product of cubed binomials)	Expanding the cube of a binomial.
20	5	Factoring	Factoring a non-perfect-square trinomial.
21	5	Factoring	Identifying a perfect-square trinomial.
22	5	Factoring	Factoring a perfect-square trinomial.
23	5	Factoring	Identifying a difference of squares (first version) / Factoring a sum of cubes (final version).
24	5	Factoring	Factoring a difference of squares.
25	5	Factoring	Factoring a difference of cubes.

Source: Own elaboration.

The design of each item is based on its respective specification, which for each item covers aspects such as the algebra topic it belongs to, its indicator according to Zabala and Arnau (2008), a comment on the meaning and functionality of the content, the foundation of the item, the vocabulary and type of information used in the item, the characteristics of the distractors, the process for obtaining the correct answer, an example item, and the estimated completion time (Table 3). The specifications preparation committee designed the specification for each item.

Table 3. Specification of item 16 on the measurement instrument.

Topic: Operations with polynomials (special product)	**Subtopic:** Squaring a binomial
Comment on the meaning of the content: Certain processes often appear in differential calculus topics, such as integral calculus and differential equations among others, and are classified as special products. This is the case with squaring a binomial which is done like this: $(a+b)^2=a^2+2ab+b^2$. It can also be written as: $(a-b)^2=a^2-2ab+b^2$. It is expected that the student will correctly square a binomial.	

Indicator of achievement: Expanding the square of a binomial.
Foundations of the item: A binomial is provided for which the students must expand its square. The binomial has the following characteristics: a) The signs of the coefficients can be the same or opposite. b) The coefficients of the terms are whole numbers. c) The literals can be any letter in the alphabet, not just *x* or *y* as is usual. d) To ensure that the item is not too difficult, use of fractional coefficients is not recommended.
Vocabulary and textual, graphic, or tabular information: The information given in this item is textual, including an algebraic expression that corresponds to the binomial and from which its square must be established.
Distractors: These will be binomials or trinomials with the following characteristics: a) The square of the first plus the square of the second. b) The square of the first minus the square of the second. c) The square of the first minus two times the product of the first and second plus the square of the second. d) The square of the first plus the first term multiplied by the second term plus the square of the second term.
Correct answer: The correct answer is obtained by following the rule for squaring a binomial: the square of the first term plus two times the product of the first term and second term plus the square of the second term.
Proposed item: Expanding the binomial $(3x-4y)^2$ gives the following result: A) $9x^2-24xy+16y^2$ B) $9x^2+16y^2$ C) $9x^2-16y^2$ D) $9x^2+24xy+16y^2$
Estimated completion time: 2 minutes.

Source: Own elaboration.

The instrument comprises 25 multiple-choice items where the student is asked to choose the correct answer from four possibilities. Each item in the instrument is independent, as it contains the information needed to approach it and answer it. The instrument is criterion-referenced as its aim is to evaluate learning and provide information about what each student can and cannot do. The items were designed by the item preparation committee.

4. Results and discussion

The pilot application of the first version of the measurement instrument was carried out in the facilities of the FIM at the UABC during the first week of the first semester of the 2018-2019 year. The instrument was applied to 177 students who had just started at the FIM and were enrolled on the Differential Calculus Course (21% of the students enrolled on this module during the semester in question). For Contreras, Backhoff, and Larrazolo (2004) and for Muñoz and Mato (2006), the reliability of the instrument calculated using the KR-20 formula r=0.88 and by the split-half method r=0.93 is classed as appropriate when it is greater than or equal to 0.85 in the case of standardised, large-scale instruments.

The mean DI was 0.70±0.19 (mean ± standard deviation). The percentage distribution for the DI is: very easy items 8% (two items), fairly easy 48% (twelve items), average difficulty 28% (seven items), fairly difficult 4% (one item), and very difficult 12% (three items).

Of the items, 68% have excellent discrimination, 24% have good discrimination, and 8% have bad discrimination. The mean IDC is 0.49±0.18 (mean ± standard deviation), which is classed as excellent.

The means of the IDC were calculated for the 5 algebra topics considered, namely: operations with fractions (items 1, 2, and 3) 0.56; law of exponents (items 4, 5, and 6) 0.61; rationalisation (items 7, 8, and 9) 0.33; operations with polynomials (items 10-19) 0.48; and factorising (items 20-25) 0.50. The highest power of discrimination was identified for the items corresponding to the law of exponents topic followed by the items from the fractions topic. In contrast, we found that the lowest power of discrimination was for the items from the rationalisation topic, which are more difficult.

The mean of the biserial correlation coefficients for the test is 0.52±0.14 (mean ± standard deviation). Of the items, 80% have excellent discrimination, 16% have good discrimination, and 4% have mediocre discrimination. No items were found with poor or worst discrimination.

Regarding content validity, the means for each item in each of the eight aspects declared in the expert evaluation were greater than or equal to 3.5. According to Contreras, Backhoff, and Larrazolo (2004), the criterion that must be satisfied is to achieve an average from the experts equal to or greater than 3.5. In this case, the instrument is classed as valid with regards to its content. Aspect 9 covered the instrument as a whole and concerned the time students are given to answer it. In this case, the maximum time permitted for completion is 60 minutes. The panel of experts believed that the time allowed for students to answer the 25 items is appropriate.

The information was extracted from the application of the differential calculus baseline test corresponding to the first semester of the 2018-2019 academic year, with answers to 60 items by 758 engineering students from the UABC's FIM being recorded. The reliability analysis using the KR-20 formula gives a coefficient $\propto$=0.87. In view of the above, the criterion used to validate the instrument is the same. Of the 177 students from the sample analysed, we have the results of 151 who sat the differential calculus baseline test. We used these to calculate the correlation between the grade obtained on the measurement instrument and the grade from the baseline test. When comparing the grades, a Pearson correlation coefficient of r=0.70 was obtained significant at the 0.01 level, classed as a significant positive correlation on the scale of Hernández, Fernández, and Baptista (2006).

The diagnostic instrument had a total of 100 options: 25 correct answers and 75 distractors. Of the distractors evaluated, 33% (25) were classified as unsuitable; 6 items (24%) had 3 suitable distractors; 14 items (56%) had 1 UD; 2 items (16%) had 2 UDs, and 1 item had 3 UDs. The mean number of UDs per item was 1.00 ± 0.74 (mean ± standard deviation).

The mean of the DI and IDC for items with 0 unsuitable distractors was 0.45 ± 0.18 and 0.46 ± 0.17, respectively (mean ± standard deviation). In contrast, these psychometric parameters in items with 2 UDs were 0.87 ± 0.03 and 0.34 ± 0.08 with 42% higher DI (in other words, 42% easier) and 12% less discrimination. For items with 1 UD, the mean DI and IDC were 0.74 and 0.58.

The ANOVA reflected statistically significant differences ($p<0.001$) in the DI between the groups of items with different numbers of UDs. The *post-hoc* Tukey test displayed significant differences ($p<0.001$) in the difficulty index (Table 4) and discrimination index between items with 0 UDs and other items with one or 2 UDs. A significant difference was found between one and 2 UDs in the IDC. With regards to the biserial correlation between the items, no significant difference was found. Item 10, which had 3 UDs, was excluded from this analysis.

Table 4. One-Way ANOVA, Difficulty Index for Groups of Items with Zero to Two Unsuitable Distractors.

Difficulty index
Tukey's B[a,b]

Unsuitable distractor	**N**	Subgroup for Alfa = .05	
		1	2
0	6	.4533	
1	14		.7450
2	4		.8725

The means for the groups in the homogenous subgroups are shown.

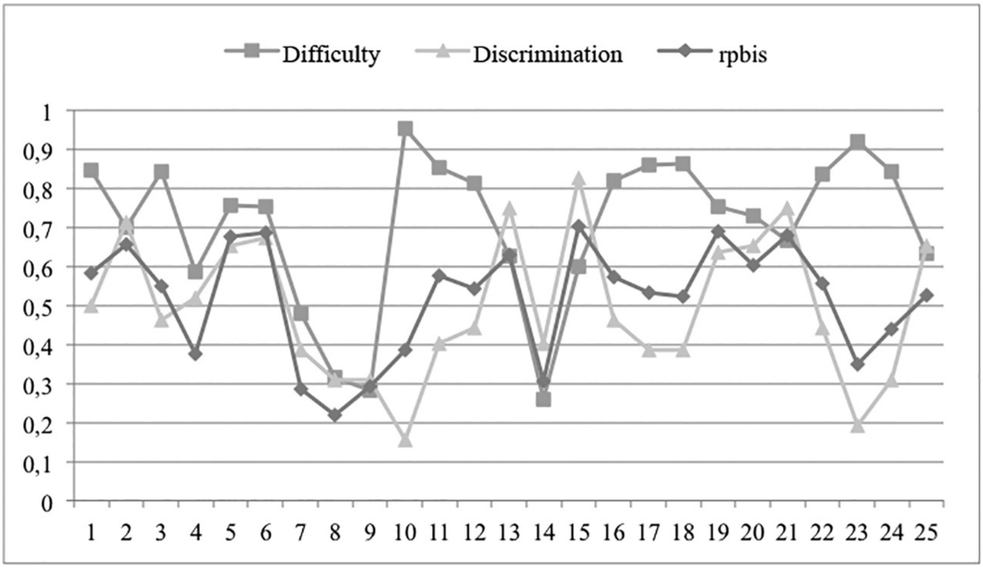
a. The sample size is used for the harmonic mean = 6.146.

b. The group sizes are not the same. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Source: Own elaboration.

The biserial correlation varies between 0.22 and 0.70 (Graph 1). A positive correlation indicates that answering the question well is an indicator of obtaining a good score on the measurement instrument. Therefore, the questions with positive rpbis are the ones that best discriminate the sample of subjects.

Graph 1. Item number compared with difficulty index, discrimination, and biserial correlation in the first version of the measurement instrument.



Source: Own elaboration.

Based on this line chart we can see that items with a difficulty greater than 0.80 (items 1, 3, 10, 11, 12, 16, 17, 18, 22, 23, and 24) have discrimination and rpbis below 0.60. In fact, the lowest discrimination values are found in this group of items. In contrast, higher discrimination values are found (classed as excellent) when the difficulty is between 0.40 and 0.75. Item 15 refers to isolating a variable and is the item with the greatest power of discrimination (IDC = 0.83). At the same time, it is the best predictor of the success of a student in this instrument in accordance with the biserial correlation coefficient (rpbis = 0.70).

To identify which algebra topics, items, and indicators of achievement lead to success for a student on this measurement instrument, we calculated the means of the biserial correlation coefficient for the algebra topics and ordered them from highest to lowest: operations with fractions (topic 1, 0.60), laws of exponents (topic 2, 0.58), operations with polynomials (topic 4, 0.55), factorising (topic 5, 0.53), and rationalisation (topic 3, 0.26). The topics that mainly predict students' success are those studied at the primary and secondary educational levels. At the individual level, the two items with the highest biserial correlation coefficient were chosen, namely: item 15, isolating a variable (rpbis = 0.70) and item 19, expanding the cube of a binomial (rpbis = 0.69).

In the two very easy items (10 and 23), the students are asked to calculate the sum of polynomials and identify a difference of squares respectively. However, given that the IDC in both cases (0.15 and 0.19) were less than 0.2 and the DI is greater than 0.9 according to the criterion of Backhoff, Larrazolo, & Rosas (2000) and CTT, we decided to modify it substantially. In the case of item 10, the expression degree was raised from first to second and the number of terms from 4 to 6. The indicator of achievement was not modified (Table 5). In the case of item 23, the topic was not changed, but the indicator of achievement was, from identifying a difference of squares to factorising a sum of cubes. Smaller changes were also made to the distractors that were not suitable.

TABLE 5. Initial and final version of item 10 from the measurement instrument.

Performing the operation 2x+7y-5x-3y gives the following result:			
A) $-3x+4y$	B) $3x-4y$	C) xy	D) $-xy$
Performing the operation $2x+7y-5x-3y-3x^2-(-7x^2)$ gives the following result:			
A) $4x^2-3x+4y$	B) $4x^2+3x-4y$	C) $-4x^2-3x+4y$	D) $-4x^2+3x+4y$

Source: Own elaboration.

After making changes to the UDs and substantially modifying items 10 and 23, we carried out a second pilot application during the first week of the second semester of 2018–2019 with a sample of 138 newly enrolled students (20% of the student body enrolled on this module during the semester in question) on the engineering degrees at the UABC's FIM. The measurement instrument was again analysed with CTT and its reliability calculated using KR-20 r=0.87 and the split-halves method r=0.92. The reliability is virtually the same with the two methods in the final version of the instrument. However, the psychometric indexes improved notably.

The information was again extracted from the application of the differential calculus baseline test. For the second semester of 2018-2019, the answers to 60 items by 627 engineering students at the FIM were recorded. The reliability analysis using the KR-20 formula gives a coefficient =0.87. Of the 138 students in the sample, we have the results of 117 students who took the differential calculus baseline test. With these results, we calculated the correlation between the grade obtained in the measurement instrument and the grade for the baseline test. Comparison of the grades gives a Pearson correlation coefficient of r= 0.72 significant at the 0.01 level, with an increase in correlation of 2.8% compared with the first version. High grades in the measurement instrument translate into high grades in the differential calculus baseline test.

In the final version of the measurement instrument, the mean DI was 0.58±0.17 (mean ± standard deviation). The percentage distribution from the DI is as follows: very easy items 4% (1 item); fairly easy 24% (6 items); average difficulty 32% (8 items); fairly difficult 36% (9 items) and very difficult 4% (1 item). This distribution best matches the criteria established by Backhoff, Larrazolo, and Rosas (2000).

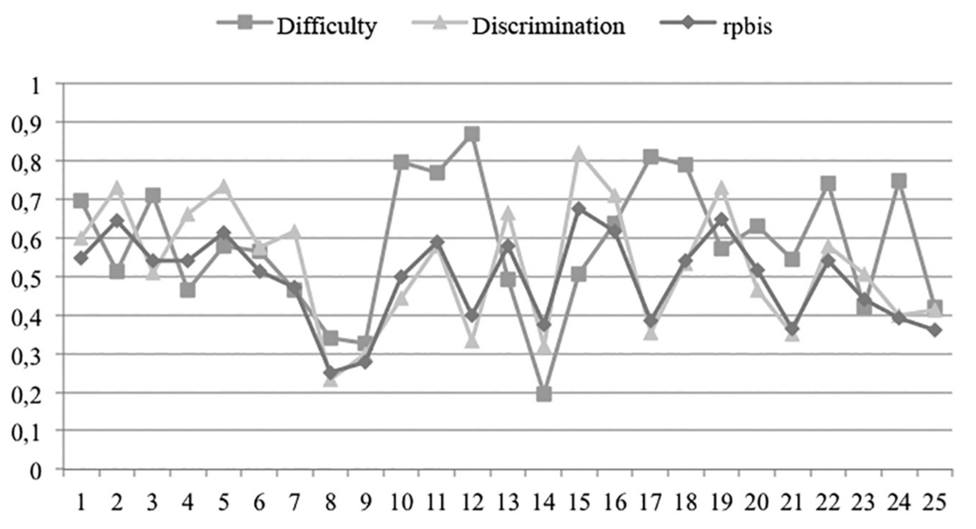
We found that 72% of the UDs have excellent IDCs and 24% have good discrimination, while 4% have mediocre discrimination. The mean IDC is 0.53±0.16 (mean ± standard deviation), which falls within the band classed as excellent. This final version of the measurement instrument increased the power of discrimination by 8%.

The mean discrimination figures were also calculated for the 5 algebra topics covered, namely: operations with fractions 0.61 (items 1, 2, and 3); law of exponents 0.65 (items 4, 5, and 6); rationalisation 0.38 (items 7, 8, and 9); operations with polynomials (items 10-19) 0.55 and factorising 0.45 (items 20-25). In the final version of the instrument, it is worth noting that we found the greatest power of discrimination in the items corresponding to the law of exponents topic, followed by the items from the fractions topic. In contrast, we found that the lowest power of discrimination still corresponds to items from the rationalisation topic, which are most difficult.

In the final version, the mean biserial correlation coefficients for each algebra topic were calculated again and were ordered from highest to lowest: operations with fractions (topic 1, rpbis = 0.58); laws of exponents (topic 2, rpbis = 0.55); operations with polynomials (topic 4, rpbis = 0.53); factorising (topic 5, rpbis = 0.44); and rationalisation (topic 3, rpbis = 0.33). Consistency was found with the topics that predict the success of students in relation to the first version of the instrument. At the particular level, the two items with the highest biserial correlation coefficients also correspond to items 15:

isolating a variable (rpbis = 0.68) and 19, cubing a binomial (rpbis = 0.65). The psychometric indicators show less dispersion in this final version of the instrument (Graph 2). The figure for the mean squared deviation from the mean for the DI is 0.90 in the first version of the instrument and 0.72 in the final version; for the IDC they are 0.79 and 0.62, and for the rpbis they are 0.51 and 0.33. The difficulty varies between 0.2 and 0.87, the discrimination between 0.23 and 0.82, and the biserial correlation coefficient varies between 0.25 and 0.68.

Graph 2. Item number against difficulty index, discrimination index, and biserial correlation in the final version of the measurement instrument.

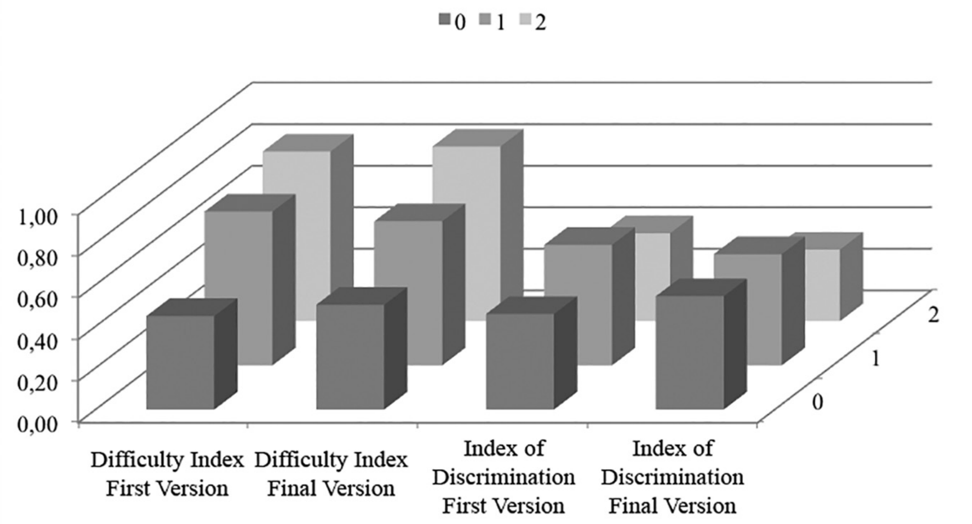


Source: Own elaboration.

Of the distractors evaluated, 15% (11) were classified as unsuitable, 16 items (64%) had 3 suitable distractors, 7 items (28%) had 1 UD, 2 items (8%) had 2 UDs, and no item had 3 UDs. The mean number of UDs per item was 0.44 ± 0.64 (mean ± standard deviation). In this final version, we reduced the UDs from 25 to 11.

The mean DI and IDC in items with 0 UDs was 0.50 ± 0.15 and 0.55 ± 0.17. In contrast, for items with 2 UDs, these psychometric parameters were 0.84 ± 0.03 and 0.34 ± 0.01 with 34% greater DI (in other words, 34% easier) and 11% less discrimination. For items with 1 UD, the mean DIs and IDCs are 0.69 ± 0.09 and 0.54 ± 0.11 respectively (mean ± standard deviation). It is confirmed that when there are 1 or 2 UDs, the item is easier (Graph 3) and the power of discrimination falls.

Graph 3. Comparison of unsuitable distractors and indexes of difficulty and discrimination in the first and final versions of the measurement instrument.



Source: Own elaboration.

A *post-hoc* Tukey test between item topics and the DI did not show significant differences between the groups. However, in the final version, the greater difficulty of answering rationalisation items (topic 3) is apparent with a mean index of 0.38, followed by the topic 2 items corresponding to the laws of exponents with a mean DI of 0.54. The next most difficult topics are those with factorising items (0.58), operations with fractions (0.64), and polynomials (0.65). Nonetheless, it is worrying that only 64% of the students starting an engineering degree can correctly solve arithmetic operations of addition, subtraction, multiplication, and division of fractions, as these skills are practised from the basic level of education.

Items 8, 9, and 14 have indexes of 0.34, 0.33, and 0.20 respectively and are the most difficult in the measurement instrument. Items 8 and 9 (fairly difficult) correspond to rationalisation of the numerator and denominator of algebraic expressions, while item 14 (very difficult) refers to calculation of the division of polynomials. Items 23 and 25 are classed as fairly difficult, both having indexes of 0.42. For these items, the student has to factor a sum of cubes and a difference of cubes respectively. This shows that the items relating to rationalisation, division of polynomials, and factorising with cubes are the most difficult ones for new students. In particular, these algebraic skills are vital to be able to calculate the limits of a function, find the zeros for an equation, or simplify algebraic expressions.

5. Conclusions

We created a valid and reliable instrument with the aim of determining the extent to which a newly enrolled student on an engineering degree has the algebra skills required to be able to take the module in Differential Calculus, and also to predict their likelihood of success on this module.

A panel of experts evaluated whether the content of the items examined the proposed algebra topics and whether the items are indicators of what is intended to be measured. The experts gave favourable opinions relating to the diagnostic potential of the measurement instrument. The differential calculus baseline test was used as the criterion to determine the criterion validity. This exam has been used at the UABC since 2005. When comparing grades from the instrument with the criterion, a Pearson correlation coefficient of 0.72 was obtained, significant at the 0.01 level. Accordingly, high grades on the measurement instrument translate into high grades on the differential calculus baseline test, and so this measurement instrument is considered to be a predictor of the students' performance on the Differential Calculus Course for engineering degrees.

We used two methods to calculate its reliability, KR-20 and split halves. The results are consistent between the first version and the final version, and so the instrument is very reliable and can be considered for large-scale use.

We found the greatest power of discrimination in the items corresponding to the topics that involve the law of exponents and operations with fractions. These topics mainly relate to the skills acquired in primary and secondary education. In contrast, we found that the lowest power of discrimination is still for the items relating to the topic of rationalisation, which the students find most difficult. We also found that the topics that predict the success of the students in the measurement instrument, in other words, the point-biserial correlation coefficient, are the operations with fractions and laws of exponents.

The results from the application of this measurement instrument make it possible to identify which algebra topics cause the most problems for students who are starting engineering degrees and will study the differential calculus module. At the same time, these results make it possible to take measures to improve the instruction and academic performance of the students on the Differential Calculus Course on the engineering programmes.

References

Alsina, Á., & 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.

Arriaga, M. (2015). El diagnóstico educativo, una importante herramienta para elevar la calidad de la educación en manos de los docentes. *Atenas, 3* (1), 63-74.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1983). *Psicología educativa: Un punto de vista cognoscitivo*. México: Editorial Trillas.

Backhoff, E., Larrazolo, N., & 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.

Carbonero, M. A., & Navarro, J. C. (2006). Entrenamiento de alumnos de educación superior en estrategias de aprendizaje en matemáticas. *Psicothema, 18* (3), 348-352.

Carmines, E., & Zeller, R. (1987). *Reliability and Validity Assessment*. USA: Sage.

Cisneros, E., Jorquera, M., & 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 curriculum, para la educación primaria en México*. Paper presented at the III Foro Nacional de Evaluación Educativa de CENEVAL, Veracruz, Mexico.

Contreras, L. A. (2000). *Desarrollo y pilotaje de un examen de español para la educación primaria en Baja California* (Doctoral dissertation, Instituto de Investigación y Desarrollo Educativo, Ensenada, México). Retrieved from https://bit.ly/2PeihuK (Consulted on 2019-10-15).

Contreras, L., Backhoff, E., & Larrazolo, N. (2004). *Educación, aprendizaje y cognición. Teoría en la práctica.* México: Manual moderno.

Contreras, L. A., Encinas, J. A., de las Fuentes, M., & Rivera, R. E. (2005). *Evaluación Colegiada del aprendizaje en la Universidad Autónoma de Baja California*. Paper presented at the VIII Congreso Nacional de Investigación Educativa of Consejo Mexicano de Investigación Educativa (COMIE), Hermosillo, México.

Corral, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencias de la Educación, 19* (33), 228-247.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* Florida: Holt, Rinehart & Winston.

Difabio, H. E. (1994). La temática de la motivación en el neoconductismo contemporáneo: locus de control y teoría de la atribución. **revista española de pedagogía***, 52* (197), 37-56.

Encinas, A., de las Fuentes, M., & Rivera, R. (2007). Construcción colegiada y aplicación de un examen criterial alineado con el currículo para evaluar a gran escala un curso de cálculo diferencial. *Acta Latinoamericana de Matemática Educativa, 20*, 192-197.

Encinas, F., Osorio, M., Ansaldo, J., & Peralta, J. (2016). El cálculo y la importancia de los conocimientos previos en su aprendizaje. *Revista de Sistemas y Gestión Educativa, 3* (7), 32-41.

García, M. B., & Vilanova, S. L. (2008). Las representaciones sobre el aprendizaje de los alumnos de profesorado. Diseño y validación de un instrumento para analizar concepciones implícitas sobre el aprendizaje en profesores de matemática en formación. *Revista electrónica de investigación en educación en ciencias, 3* (2), 27-35.

Gatica-Lara, F., Méndez-Ramírez, I., Sánchez-Mendiola, M., & Martínez-González, A. (2010). Variables asociadas al éxito académico de los estudiantes de la Licenciatura en Medicina de la UNAM. *Revista de la Facultad de Medicina de la UNAM, 53* (5), 9-18.

González, E. I. (2013). Estudio sobre factores contexto en estudiantes universitarios para conocer por qué unos tienen éxito mientras otros fracasan. *Revista Intercontinental de Psicología y Educación, 155* (2), 135-154.

González, R. M. (2005). Un modelo explicativo del interés hacia las matemáticas de las y los estudiantes de secundaria. *Educación Matemática, 17* (1), 107-128.

Guilford, J. P. (1975). *Psycometric Methods*. Nueva Delhi: McGraw-Hill.

Hernández, R., Fernández, C., & Baptista, P. (2006). *Metodología de la investigación*. México: Mc Graw-Hill.

Henrysson, S. (1971). Gathering, Analysing, and Using Data on Teste Items. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 130-159). Washington: American Council on Education.

Morales, E. M. (2009). Los conocimientos previos y su importancia para la comprensión del lenguaje matemático en la educación superior. *Universidad, Ciencia y Tecnología, 13* (52), 211-222.

Muñoz, J. M., & Mato, M. D. (2006). Diseño y validación de un cuestionario para medir las actitudes hacia las matemáticas de los alumnos de ESO. *Revista Galego-Portuguesa de Psicología e Educación, 13* (11-12), 413-424.

Nitko, A. J. (1994). *A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students*. Paper presented at the Conferencia Internacional sobre Evaluación y Medición Educativas of the Asociación para el Estudio de la Evaluación Educativa (ASSESA), Pretoria, South África.

Orozco, C., & Díaz, M. A. (2009). Atribuciones de la motivación al logro y sus implicaciones en la formación del pensamiento lógico-matemático en la universidad. *Interciencia, 34* (9), 630-636.

Orozco-Moret, C., & Morales, V. (2007). Algunas alternativas didácticas y sus implicaciones en el aprendizaje de contenidos de la teoría de conjuntos. *Revista Electrónica de Investigación Educativa, 9* (1), 1-19.

Prieto, G., & Delgado, A. R. (2010). Fiabilidad y validez. *Papeles del Psicólogo, 3* (1), 67-74.

Popham, J. (1990). *Modern educational Measurement: a practitioner's perspective.* Boston: Prentice Hall.

Reidl-Martínez, L. M. (2013). Confiabilidad en la medición. I*nvestigación en Educación Médica, 2* (6), 107-111.

Reynoso, O., & Méndez-Luévano, T. E. (2018). ¿Es posible predecir el rendimiento académico? La regulación de la conducta como un indicador del rendimiento académico en estudiantes de educación superior. *Diálogos sobre educación. Temas actuales en investigación educativa, 9* (16), 1-19.

Rodríguez, O., Casas, P., & Medina, Y. (2005). Análisis Psicométrico de los exámenes de evaluación de la calidad de la educación superior (ECAES) en Colombia. *Avances en Medición, 3*, 153-172.

Sells, L. W. (1973). *High School Mathematics as the Critical Filter in the Job Market*. Retrieved from https://bit.ly/33YD5uy (Consulted on 2019-10-15).

SEP (2017). *Planes de estudio de referencia del componente básico del marco curricular común de la educación media superior.* México: Secretaría de Educación Pública.

Zabala, A., & Arnau, L. (2008). *11 Ideas clave, ¿Cómo aprender y enseñar competencias?* Barcelona: Editorial Grao.

Authors' biographies

Wendolyn Elizabeth Aguilar Salinas holds a Doctorate in Sciences from the Universidad Autónoma de Baja California. She is currently Associate Professor and Head of Tutoring at the Mexicali Faculty of Engineering of that university. Her research focusses on the learning and teaching of mathematics, on the teaching techniques and technology, and on educational modalities.

http://orcid.org/0000-0003-2223-9234

Maximiliano de las Fuentes Lara holds a Doctorate in Higher Education in Engineering from the Universidad Autónoma de Baja California. He is currently Associate Professor and Coordinator of the differential and integral calculus modules at the Mexicali Faculty of Engineering at that university. His research focusses on the problem of teaching, learning, and evaluation of mathematics for engineering.

http://orcid.org/0000-0002-1001-4663

Araceli Celina Justo López holds a Doctorate in Engineering from the Universidad Autónoma de Baja California. She is currently Associate Professor and Coordi-

nator of the common core at the Mexicali Faculty of Engineering at that university. Her research focusses on educational technologies.

http://orcid.org/0000-0002-6911-2065

Ana Dolores Martínez Molina holds a Master's in Network and Information Technology at CETYS Universidad. She is currently Associate Professor and Head of Academic Guidance at the Universidad Autónoma de Baja California. Her research interests focus on mathematics in education.

https://orcid.org/0000-0001-5130-6243

revista española de pedagogía
año 78, nº 275, enero-abril 2020
Spanish Journal of Pedagogy
year 78, n. 275, January-April 2020

Table of Contents

Sumario

This is the English version of the research articles and book reviews published originally in the Spanish printed version of issue 275 of the **revista española de pedagogía**. The full Spanish version of this issue can also be found on the journal's website http://revistadepedagogia.org.

ISSN: 0034-9461 (Print), 2174-0909 (Online)
https://revistadepedagogia.org/
Depósito legal: M. 6.020 - 1958
INDUSTRIA GRÁFICA ANZOS, S.L. Fuenlabrada - Madrid