Revista Española
de Educación Médica

Design and validation of an instrument to evaluate intrinsic and extrinsic factors of academic performance in medical students.

Diseño y validación de un instrumento para evaluar factores intrínsecos y extrínsecos del rendimiento académico en estudiantes de medicina.

Gustavo Domínguez-Celis[1], Lorena Mendoza-Anaya[2], Sandra Alicia Reza-López[3], María Cecilia Ishida-Gutiérrez[4], Omar Fierro-Fierro [5]

[1] Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua; drgustavodomc@gmail.com
[2] Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua; lmendozaa@uach.mx ; https://orcid.org/0000-0003-1302-8785
[3] Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua; sreza@uach.mx ; https://orcid.org/ 0000-0001-5541-5308
[4] Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua; mishida@uach.mx ; https://orcid.org/0000-0002-7642-2328
[5] Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua; fierro_omar@hotmail.com

* Correspondence: lmendozaa@uach.mx

Received: 13/1/26; Accepted: 4/2/26; Published: 6/2/26

Summary.

Background: Low academic performance is a significant problem in medical programs, affecting students' educational trajectory and well-being. This phenomenon results from the complex interaction between intrinsic (cognitive and emotional) and extrinsic (contextual and social) factors. However, there is a persistent lack of validated instruments that comprehensively assess this multiplicity of factors in the Mexican context. The systematic identification of these factors is essential for designing effective preventive interventions. **Objective:** To design and validate an instrument to systematically identify and assess the intrinsic and extrinsic factors that influence the academic performance of medical students. **Method:** An instrumental study with an exploratory sequential design. Phase I: a literature review in six specialized databases, 57 exploratory interviews with discourse analysis, and preliminary instrument development. Phase II: two pilot tests (n_1=33, n_2=45) and two rounds of validation by expert judgment (5 and 6 experts, respectively). The Content Validity Coefficient (CVC) was used as a measure of reliability. **Results:** The first round of validation of the 42 preliminary items yielded a total CVC of 0.902. After incorporating expert recommendations, the final 38-item version achieved a total CVC of 0.974, exceeding international acceptance criteria. The final instrument integrates four dimensions that operationalize the key factors: (1) sociodemographic characteristics and academic background; (2) metacognition and self-regulation; (3) academic emotions and support network; and (4) sleep habits and mental health. **Conclusions:** A validated, reliable, and comprehensive instrument was developed that operationalizes the intrinsic (metacognition, emotional regulation, mental health) and extrinsic (academic background, socioeconomic context, support network) factors that determine academic performance in Mexican medical students. This tool facilitates the early identification of academic

risk and provides a basis for the design of contextualized and personalized educational interventions.

Keywords: Academic performance, intrinsic-extrinsic factors, medical students, instrument validation.

Resumen.

Antecedentes: El bajo rendimiento académico es una problemática significativa en los programas de medicina que afecta la trayectoria formativa y el bienestar estudiantil. Este fenómeno resulta de la interacción compleja entre factores intrínsecos (cognitivos y emocionales) y extrínsecos (contextuales y sociales). Sin embargo, persiste una carencia de instrumentos validados que evalúen de forma integral esta multiplicidad de factores en el contexto mexicano. La identificación sistemática de estos factores es esencial para diseñar intervenciones preventivas efectivas. **Objetivo:** Diseñar y validar un instrumento para identificar y evaluar de manera sistemática los factores intrínsecos y extrínsecos que influyen en el rendimiento académico de estudiantes de medicina. **Método:** Estudio instrumental con diseño secuencial exploratorio. Fase I: revisión de literatura en seis bases de datos especializadas, 57 entrevistas exploratorias con análisis de discurso, y construcción preliminar. Fase II: dos pruebas piloto (n_1=33, n_2=45) y dos rondas de validación por juicio de expertos (5 y 6 expertos respectivamente). Se empleó el Coeficiente de Validez de Contenido (CVC) como medida de confiabilidad. **Resultados:** La primera ronda de validación de los 42 ítems preliminares obtuvo un CVC total de 0.902. Tras incorporar las recomendaciones de los expertos, la versión final de 38 ítems alcanzó un CVC total de 0.974, superando los criterios de aceptación internacionales. El instrumento final integra cuatro dimensiones que operacionalizan los factores clave: (1) características sociodemográficas y antecedentes académicos; (2) metacognición y autorregulación; (3) emociones académicas y red de apoyo; y (4) hábitos de sueño y salud mental. **Conclusiones:** Se generó un instrumento validado, confiable e integral que operacionaliza los factores intrínsecos (metacognición, regulación emocional, salud mental) y extrínsecos (antecedentes académicos, contexto socioeconómico, red de apoyo) determinantes del rendimiento académico en estudiantes de medicina mexicanos. Esta herramienta facilita la identificación temprana de riesgo académico y fundamenta el diseño de intervenciones educativas contextualizadas y personalizadas.

Palabras clave: Rendimiento académico, factores intrínsecos-extrínsecos, estudiantes de medicina, validación de instrumentos.

---

1. Introduction

Academic performance can be understood as the set of actions a student undertakes to achieve their academic goals (1). Academic performance encompasses both quantitative and qualitative dimensions, reflected in the profile of skills, knowledge, attitudes, and values developed at the end of the teaching-learning process (2). In medical education, academic performance is a key indicator of educational success and institutional efficiency, and it is especially relevant considering that, although medical degrees have relatively lower dropout rates compared to other disciplines, each student who drops out represents a significant loss of institutional resources, human capital, and opportunities for future healthcare (3-4). Recent national literature consistently documents that medical and health sciences programs show the lowest dropout rates compared to other degrees; for example, the Faculty of Medicine at the National Autonomous University of Mexico (UNAM) reports dropout rates ranging from 5% to 17% (4-6). These figures contrast sharply with general estimates for higher education in Mexico, where dropout rates can approach 40%, according to reports analyzed in recent educational research (7). Furthermore, it has been noted that low

academic performance is a significant contributing factor to the desire to drop out of school, with up to 11% of students citing this as a determining factor (3).

Academic performance is a multifactorial phenomenon, influenced by a complex interaction of intrinsic student factors, such as mental health, motivation, and study habits, which have been shown to be significantly related to academic achievement (8-10); and extrinsic factors, ranging from family support and socioeconomic environment to teaching quality and the educational environment (10-14). In medical education, factors associated with academic failure include inadequate planning and self-regulation, psychological and personality problems, sociodemographic factors, cognitive problems, demotivation, low self-efficacy, a poor learning environment, and inappropriate teaching strategies (15-16). Recent specialized research in medical education has revealed alarming prevalence rates of academic stress; up to 69.62% of medical students experience severe stress, 26.80% moderate stress, and 3.58% mild stress. Overwork and constant pressure in medical training can lead to the development of mental health symptoms, such as *burnout syndrome,* which has an incidence of 33% to 55%, as well as depression and anxiety. This directly affects students' learning capacity and academic performance (17-23).

However, this favorable national trend is not uniform across all institutions. Research conducted by Reyes Seáñez et al. (24) with medical students at the Faculty of Medicine and Biomedical Sciences (FMyCB) of the Autonomous University of Chihuahua (UACH) identifies a specific institutional situation that requires attention: they observed that up to 50% of first-semester students fail to enroll as regular students in the second semester because they decided not to continue their studies of their own volition, because they were expelled from the academic system after failing three or more courses, or because they had to retake the failed courses. The authors point out that the Human Anatomy course critically reflects this problem, with a failure rate of 61% in the first semester. Although this figure represents an atypical situation compared to the national averages reported in other Mexican medical schools (4-6), it demonstrates the existence of particular institutional contexts where the problem of low academic performance and student underachievement is especially relevant. This situation negatively impacts the student's academic trajectory and university integration, conditioning their future opportunities, such as the choice of locations for undergraduate internships and access to specialty positions.

Despite the evidence surrounding this problem, a lack of instruments that allow for a comprehensive and contextualized analysis of this phenomenon has been identified. While internationally validated instruments exist to measure constructs such as motivation, stress, or learning strategies in isolation, research consistently reports the limitations of these partial approaches (25-27). This gap is particularly evident in low- and middle-income countries such as Mexico. In this context, Barteit et al. (28) have reported that most e-learning studies *for* medical education have used customized and unvalidated instruments, resulting in low comparability and poor validity. This highlights the need for a holistic instrument, designed and validated for the medical student population in the Mexican context. Even in contexts where dropout rates in medicine are relatively low compared to other disciplines, each student who leaves the program represents not only a loss of institutional resources invested in their training but also a reduction in the human capital available to address the population's health needs. In the Mexican context, characterized by institutional heterogeneity where programs with favorable indicators coexist (4-6) with others facing specific challenges related to academic performance (24), the development of holistic and contextualized diagnostic instruments acquires strategic relevance for both prevention and timely intervention. Mexico faces a complex reality in medical training, with an explosive growth in the number of medical schools, which has increased from a few institutions to more than 160 educational programs in recent decades. This expansion has generated concerns about

educational quality, especially considering that only 50% of the country's medical schools are accredited by the Mexican Council for the Accreditation of Medical Education (COMAEM) (29-30).

In this context of institutional heterogeneity and variability, the early identification of risk factors using specialized instruments allows for the implementation of support strategies before students consider dropping out, thus maximizing retention rates and academic success. Therefore, the objective of this study was to design and validate an instrument to systematically identify and evaluate the intrinsic (personal) and extrinsic (contextual) factors that influence the academic performance of medical students in the Mexican context.

2. Methods

The research was developed at the facilities of the Faculty of Medicine and Biomedical Sciences of the Autonomous University of Chihuahua (FMyCB of the UACH).

2.1 Study Design

The research was instrumental with an exploratory sequential design. It was divided into two phases:

2.2 Phase I: Qualitative Exploration and Preliminary Construction, consisted of a literature review in specialized databases (PubMed, Scopus, Scielo, Web of Science, Elsevier and Dialnet) carried out between January and July 2024, using terms related to academic performance, intrinsic/extrinsic factors and medical students. The methodological process integrated qualitative discourse analysis (DA) techniques with quantitative psychometric validation methods, recognizing student discourse as a social action practice linked to specific institutional conditions (31-33). A non-probability convenience sampling method was used. The sample consisted of two groups of students from the selective semester enrolled in the "Life Skills Seminar" during the period August-December 2024: the first group comprised 22 students and the second 35 students. Both groups were administered two ad hoc questionnaires for contextual exploration. The sample size for these exploratory interviews (n=22 and n=35) was determined following qualitative research principles, where the sufficiency criterion is based on the theoretical saturation of emergent categories through discourse analysis, rather than on statistical representativeness, in accordance with the methodological principles established for feasibility studies (34). The responses provided in the questionnaires were treated as discursive units representative of social constructions of educational experience. Specific discourse analysis strategies were applied: analysis of multivocality, lexical content, and contextual dimensions, following the principle of "analytical vigilance" to identify emerging categories (31). Through interpretive synthesis of bibliographic findings and discourse analysis, an initial version of 49 open-ended items organized into five dimensions was developed.

2.3 Phase II: Pilot Tests, Content Validation and Adjustments.

2.3.1 First pilot test. The preliminary instrument in physical format was administered to 33 first-semester students using non-probability sampling. The participants had an average age of 20.12 years; approximately one-third were male, one-third came from other cities (students who moved to Chihuahua to study), 18.18% depended at least partially on family members to cover their school expenses, 39.39% worked at least occasionally, and 78.79% had obtained a grade point average of 8.5 or higher upon graduating from high school. The sample size of 33 students meets contemporary methodological recommendations that suggest between 30 and 50 participants for pilot studies intended to evaluate the clarity, comprehensibility, and application time of instruments (34-36), far exceeding the minimums established in recent literature (12-30 participants) (37-39). This pilot test allowed us to identify redundant elements, questions that were difficult to

understand, and to optimize the instrument's completion time, reducing the average time from 54 minutes to a more manageable duration. Based on these findings, modifications were made to the instrument, resulting in a refined version of 42 items (reduced from the initial 49 items).

2.3.2. Content Validation by Expert Judgment. First round of evaluation : a panel of five expert judges was formed, selected based on specific criteria of experience in teaching, research, education, psychiatry, and neuroscience applied to the educational context (Table 1). This panel meets the updated methodological criteria established in recent specialized literature, which recommends between 3 and 10 experts, with 5-7 being the optimal range for content validation of research instruments (40-43). A panel of this size provides sufficient control over random agreement while maintaining efficiency in the validation process, since increments greater than 10 experts do not contribute significant differences (40). Furthermore, selecting an odd number of judges minimizes the probability of ties in the evaluations, facilitating consensus formation (44). Each expert evaluated the 42 items using four Likert-type scales (1-4), where 1 indicated no rating, 2 low, 3 medium, and 4 high, according to the methodological criteria of Escobar Pérez and Cuervo Martínez (45) and Carrera Farran et al. (46). The aspects evaluated were:

- Clarity: Syntactic and semantic evaluation to ensure straightforward understanding.
- Coherence: Logical consistency between items in the set.
- Relevance: Ability to capture relevant experiences from the studied context.
- Unambiguity: An unequivocal formulation that avoids interpretive ambiguities.

2.3.3 Quantitative and qualitative analysis of the first round of evaluation. The Content Validity Coefficient (CVC) was calculated according to Hernández Nieto (47), using a Microsoft Excel macro designed by Maldonado Suárez and Santoyo Telles (48). The CVC measures the degree of agreement or concordance among experts regarding the content validity of a measurement instrument. Its calculation is performed in five stages: (1) obtaining the arithmetic mean of the scores given by the experts to the item (univocity, clarity, coherence, and relevance) using Mx = Σ(scores)/#Judges; (2) calculating the CVCi by dividing the mean by the maximum value of the scale, that is, the maximum possible score on the scale (for example, 4 if using a 1-4 scale or 5 if using a 1-5 scale), so the maximum achievable value would be 16 in this case because four criteria are being used with a maximum possible score of four. (CVCi = Mx/Vmax); (3) estimate the probability of error using Pei = (1/Vmax)^#Judges; (4) correct the CVCi by subtracting the error to obtain the final item coefficient (CVCc = CVCi - Pei); and (5) average all the CVCcs to obtain the total instrument coefficient (CVCt = Σ CVCc/n). Hernández Nieto (47) suggests acceptance criteria of CVCic > 0.80 for individual items and CVCt > 0.90 for the overall instrument. In the qualitative review of the observations and suggestions by expert judges, specific problems were identified that required the elimination of 4 items, resulting in an instrument with 38 items.

2.4 Second round of validation. A new independent panel of six expert judges (Table 2) evaluated the refined 38-item instrument. This composition exceeds the recommended minimum of 5 experts and is within the optimal range established in recent instrument validation studies in medical education, thus strengthening the robustness of the content validation (43, 49). The use of an independent panel in the second round minimizes validation bias and provides an impartial evaluation of the refined instrument, following best methodological practices in contemporary instrumental research (41-42).

2.4.1. Quantitative and qualitative analysis of the second round of evaluation. The Content Validity Coefficient (CVC) was calculated as described above for the evaluation conducted by the six expert judges. A qualitative review of the observations and suggestions by the expert judges was conducted, and it was not necessary to modify the number of items or the fundamental structure of any of them. Second pilot test: The refined instrument was administered via Google Forms to 45

first-semester students with an average age of 19.78 years. The demographic composition was as follows: 64.44% female and 35.56% male; approximately one-third came from other cities (outsiders), 8.89% were only partially dependent on their parents for expenses, 20% worked at least occasionally, 95.56% had parents with at least a high school education, and 77.78% had obtained a grade point average of 8.5 or higher upon graduating from high school. The sample size of 45 students falls within the upper limit of the optimal range recommended by recent methodological literature (30-50 participants) (34-36), allowing for solid evidence regarding the comprehensibility, acceptability, and performance of the instrument in its final 38-item version. This size significantly exceeds the minimum of 30 participants suggested for evaluating questionnaire reliability considering non-response rates (38), confirming an acceptable completion time of 37 minutes and the collection of relevant responses that provided the expected information.

Table 1. Sociodemographic and educational characteristics of the first group of experts.

Category	Description	Frequency	Percentage
Sex	Male	2	40
	Female	3	60
Highest level of education	Mastery	2	40
	Doctorate	3	60
Professional experience in research (years)	0	1	20
	1-10	2	40
	10-20	0	0
	>20	2	40
Number of articles published	1-10	2	40
	11-20	0	0
	>20	3	60
Number of book chapters published	1-2	3	60
	3-4	1	20
	5-6	1	20
Number of books published	0	4	80
	1-5	0	0
	6-10	0	0
	11-15	1	20
Professional experience in teaching	Yeah	5	100
	No	0	0

Table 2. Sociodemographic and educational characteristics of the second group of experts.

Category	Description	Frequency	Percentage
Sex	Male	5	83.3
	Female	1	16.7
Highest level of education	Mastery	0	0.0
	Doctorate	6	100.0
Professional experience in research (years)	0	0	0.0
	1-10	3	50.0
	10-20	3	50.0
	>20	0	0.0
Number of articles published	1-10	2	33.3
	11-20	2	33.3
	>20	2	33.3
Number of book chapters published	1-10	3	50.0
	11-20	1	16.7
	>20	2	33.3
Number of books published	0	1	16.7
	1-3	3	50.0
	4-6	1	16.7
	7-9	1	16.7
Professional experience in teaching	Yeah	6	100.0
	No	0	0.0

3. Results

3.1 Phase I: Qualitative Exploration and Preliminary Construction

The first exploratory phase included 22 students from the selective semester, predominantly female (68.2%), with an average age of 19.3 years, as well as a second sample of 35 students (14 previous participants and 21 new ones), maintaining a similar distribution by gender (62.9% female) and average age (19.8 years).

3.2. Discourse Analysis and Construction of the Preliminary Instrument

The discourse analysis of the exploratory responses and the literature review resulted in a preliminary instrument of 49 items structured in five dimensions: (1) Sociodemographic characteristics and academic background (6 items), (2) Metacognition (9 items), (3) Academic emotions subdivided into emotional management and support network (19 items), (4) Feelings, thoughts and mental health (7 items), and (5) Sleep habits and hygiene (8 items).

3.3 Phase II. Pilot Tests, Content Validation and Adjustments. First pilot test

The first pilot test, with 33 students, identified redundant elements and difficult-to-understand questions, as well as an average completion time of 54 minutes, leading to the elimination of 7 items. This structural optimization process resulted in a refined 42-item instrument with greater contextual relevance and a more specific population focus.

3.3.1 First Round of Content Validation by Expert Judgment

The first panel of five expert judges evaluated the refined 42-item instrument as shown in Table 3. The results revealed an overall CVCt of 0.902, where 23 items scored CVCic ≥ 0.90 (Table 4), 14 items scored CVCic > 0.80, and five items required modifications due to CVCic ≤ 0.80.

3.3.2 Qualitative review

Through a review of the feedback on the 42 items evaluated by the five expert judges, five main areas for improvement were identified: question structure (17 comments), clarity of wording (10), contextual specificity (10), response format (7), and thematic relevance (1). The resulting modifications included the elimination of four items due to redundancy and lack of clarity, as well as the modification of five items, resulting in a 38-item instrument.

Table 3. Evaluation of the preliminary instrument by a panel of five expert judges.

Item	Univocity	Relevance	Coherence	Clarity	Σ	CVCic
1	18	19	18	16	71	0.887
2	20	20	20	20	80	1
3	20	20	20	20	80	1
4	16	18	18	17	69	0.862
5	17	18	18	17	70	0.875
6	17	18	18	17	70	0.875
7	19	19	19	19	76	0.95
8	18	18	17	16	69	0.862
9	18	18	18	18	72	0.9
10	18	20	19	20	77	0.962
11	16	17	17	17	67	0.837
12	15	17	17	16	65	0.812
13	18	18	18	18	72	0.9
14	20	20	20	20	80	1
15	19	19	19	18	75	0.937
16	15	17	19	16	67	0.837
17	16	17	18	17	68	0.85
18	15	17	18	17	67	0.837
19	13	16	16	17	62	0.775
20	15	16	16	17	64	0.8
21	17	17	17	16	67	0.837
22	20	19	19	20	78	0.975
23	16	16	17	16	65	0.812
24	17	20	20	20	77	0.962
25	19	19	19	19	76	0.95
26	20	20	20	20	80	1
27	20	20	20	19	79	0.987
28	19	17	19	19	74	0.925
29	17	19	19	18	73	0.912
30	18	17	19	17	71	0.887
31	14	16	16	13	59	0.737
32	14	15	17	15	61	0.762
33	17	18	18	16	69	0.862
34	14	16	16	14	60	0.75
35	19	20	20	18	77	0.962
36	19	20	20	20	79	0.987
37	18	20	19	18	75	0.937
38	19	20	20	20	79	0.987
39	18	18	18	18	72	0.9
40	20	20	20	20	80	1

41	19	20	20	20	79	0.987
42	20	20	20	20	80	1
			$CVC_{total} = 0.902$			

Table 4. Content validation coefficient of the preliminary version of the instrument.

Dimension	Number of items	Number and % of Items with an "Excellent" validation coefficient	CVC range
Dimension 1	9	4 (44%)	0.862-1
Dimension 2	6	4 (67%)	0.812-1
Dimension 3	3	0 (0%)	0.837-0.85
Subdimension 1	5	1 (20%)	0.775-0.975
Subdimension 2	6	6 (100%)	0.912-1
Dimension 4	5	0 (0%)	0.737-0.887
Dimension 5	8	8 (100%)	0.937-1
Total	**42**	**23 (55%)**	**0.737-1**

3.3.3 Second Round of Content Validation by Expert Judgment

The second independent panel of six expert judges evaluated the refined 38-item instrument, as shown in Table 5. The results showed a significant improvement, with an overall CVCt of 0.974. Furthermore, 36 items achieved a CVCic ≥ 0.90, 2 items a CVCic > 0.80, and no item had a CVCic ≤ 0.80. This significantly exceeded both the minimum criterion of 0.80 and the CVCt of the previous 42-item instrument (0.902). The review of the expert judges' comments revealed a notable reduction in the number of areas for improvement, with only 19 comments distributed across: question structure (9), contextual specificity (4), response format (3), clarity of wording (2), and thematic relevance (2). This reduction indicated greater expert consensus regarding the quality of the instrument.

3.4 Second pilot test

The second pilot test with 45 students confirmed the instrument's excellent acceptability and comprehension. The average completion time was 37 minutes, and the responses obtained were relevant and provided the expected information.

3.5 Final Instrument

The process described in the previous phases resulted in an instrument consisting of 38 items, distributed in four main dimensions (table 7): (1) Sociodemographic characteristics and academic background (7 items), (2) Metacognition (13 items), (3) Academic emotions, emotional management and support network (10 items), and (4) Sleep habits, mental health and general well-being (8 items).

3.5.1 Validity Indicators by Dimension

The final instrument obtained an overall CVCt of 0.974. All dimensions of the final instrument obtained content validity indices greater than 0.80, with 36 of 38 items (94.74%) achieving the "Excellent" classification (Table 6) according to the criteria proposed by Hernández Nieto (47), demonstrating relevance and univocity to identify intrinsic and extrinsic factors that influence academic performance.

Table 5. Evaluation of the refined instrument by a panel of six expert judges.

Item	Univocity	Relevance	Coherence	Clarity	Σ	CVCic
1	23	23	24	24	94	0.979
2	23	22	24	24	93	0.969
3	24	24	24	24	96	1,000
4	21	23	23	21	88	0.917
5	22	24	24	24	94	0.979
6	24	24	24	24	96	1,000
7	22	23	23	24	92	0.958
8	19	22	23	23	87	0.906
9	22	23	24	24	93	0.969
10	22	23	23	23	91	0.948
11	19	23	21	21	84	0.875
12	23	24	24	24	95	0.990
13	24	24	24	24	96	1,000
14	23	24	24	24	95	0.990
15	23	24	24	24	95	0.990
16	23	24	24	24	95	0.990
17	22	23	24	23	92	0.958
18	17	20	21	22	80	0.833
19	23	24	24	24	95	0.990
20	23	24	23	24	94	0.979
21	23	23	24	24	94	0.979
22	22	24	24	24	94	0.979
23	22	24	24	24	94	0.979
24	24	24	24	24	96	1,000
25	23	24	24	23	94	0.979
26	24	24	24	24	96	1,000
27	23	24	24	24	95	0.990
28	23	24	24	24	95	0.990
29	24	24	24	24	96	1,000
30	23	24	24	24	95	0.990
31	24	24	24	24	96	1,000
32	24	24	24	24	96	1,000
33	24	24	24	24	96	1,000
34	22	23	24	24	93	0.969
35	24	24	24	24	96	1,000
36	24	24	24	24	96	1,000
37	22	24	24	24	94	0.979
38	22	23	24	24	93	0.969
$CVC_{total} = 0.974$						

Table 6. Content validation coefficient of the final version of the instrument.

Dimension	Number of items	Number and % of Items with an "Excellent" validation coefficient	CVC range
Dimension 1	7	7 (100%)	0.917-1
Dimension 2	13	11 (84.62%)	0.833-1
Dimension 3	10	10 (100%)	0.979-1
Dimension 4	8	8 (100%)	0.969-1

Total	**38**	**36 (94.74%)**	**0.833-1**

Table 7. Dimensions that make up the final instrument.

Dimension 1: Sociodemographic Characteristics and Academic Background	This dimension assesses the student's entry factors, including demographic variables, previous academic background, family socioeconomic context, and previous educational trajectory.
Dimension 2: Metacognition and Learning Strategies	It focuses on measuring the student's self-regulatory abilities and independent learning skills. Specifically, it assesses the student's capacity to plan their studies, monitor their learning progress, identify their academic strengths and weaknesses, and adjust their study strategies according to the demands of each subject.
Dimension 3: Academic Emotions, Emotional Management and Support Network	It measures the emotional and social component of academic performance, recognizing that medical learning occurs in an emotionally demanding context. It assesses the emotions students experience before, during, and after critical academic situations such as exams, presentations, and interactions with instructors.
Dimension 4: Sleep Habits, Mental Health and General Wellbeing	It measures sleep patterns, sleep quality, sleep hygiene habits, and how these relate to the ability to concentrate and retain complex medical information. It assesses mental health indicators such as symptoms of depression, anxiety, chronic stress, and academic burnout. It also includes measurements of self-care habits, work-life balance, leisure activities, and stress management strategies.

4. Discussion

A 38-item instrument was developed that has demonstrated excellent content validity for assessing the factors that influence the academic performance of medical students. This result is the culmination of a rigorous and transparent iterative validation process that combined quantitative and qualitative analyses. The final version of the instrument achieved a CVCt of 0.974. According to Hernández Nieto's interpretive framework (47), a value greater than 0.90 is classified as "Excellent." The methodological process demonstrates how qualitative feedback and empirical testing can transform a good instrument into an excellent one, ensuring that each item possesses high clarity, coherence, relevance, and unambiguity, as evidenced by the differences between the first and second versions of the instrument. These findings are consistent with the validation frameworks of Escobar Pérez and Cuervo Martínez (45) and Carrera Farran et al. (46). Methodological rigor is a fundamental pillar in educational research, where expert validation is considered a cornerstone to guarantee the quality of measurement instruments.

4.1 Diagnostic Utility and Institutional Benefits

The instrument designed and validated in this research comprehensively and systematically measures intrinsic factors of students, such as resilience, motivation, and study techniques and habits, among others; and extrinsic factors, such as the family, academic, and social context, that influence the academic performance of medical students. Its central construct is based on a multifactorial model of academic performance that recognizes the complex interaction between students' personal variables and contextual elements of the educational environment. Unlike partial instruments that evaluate isolated constructs such as motivation, stress, or learning strategies, this instrument provides a holistic assessment that allows for the systematic identification of the set of

factors that impact the academic experience of medical students. The instrument is structured around four main dimensions, each of which measures specific aspects of the phenomenon of academic performance and allows for the identification of academic risk patterns before they manifest as failing grades or dropping out. This diagnostic capacity allows for the classification of students according to their risk profile and specific intervention needs.

At the student level, the instrument offers unique early diagnostic capabilities that allow for the identification of risk patterns from the first semester, before they manifest as tangible academic failures. This preventative detection contrasts significantly with traditional indicators that evaluate performance after assessments. Open-ended responses generate individualized profiles that reveal not only which factors are present, but also how they interact in each specific case, enabling the design of highly targeted interventions. For example, when thematic analysis identifies metacognitive deficiencies in planning and study strategies, specific workshops on effective learning techniques, such as those described by Bin Abdulrahman et al. (8), can be implemented. Similarly, emotional patterns that demonstrate demotivation or anxiety justify specialized student well-being programs, ranging from *mindfulness -based interventions* to individualized psychological intervention systems.

At the teacher and institutional levels, the instrument functions as a multidimensional approach that examines both the overt learning environment and the hidden curriculum. The anonymous analysis of student perceptions of classroom interactions, pedagogical clarity, and support received provides direct, constructive feedback that guides teacher training programs focused on safer, more inclusive, and more effective learning environments. Simultaneously, by investigating emotions, sources of stress, and perceived power dynamics, the instrument reveals problematic aspects of the hidden curriculum, such as excessive competition or the normalization of mistreatment, making the invisible visible to promote institutional reflection that aligns the formal curriculum with values of professionalism, empathy, and resilience.

4.2 Systemic and Accreditation Benefits

The systematic implementation of this instrument generates benefits that extend to compliance with national and international accreditation standards in medical education. Organizations such as the Mexican Council for the Accreditation of Medical Education (COMAEM) and the *World Federation for Medical Education* require evidence of quality assurance systems and continuous improvement processes based on data (50). This instrument not only produces this evidence but also generates valuable longitudinal data for institutional educational research, enabling cohort studies that identify temporal trends, seasonal patterns, and the effects of specific interventions on academic performance. This capacity for ongoing research contributes to the development of a robust institutional evidence base that grounds educational decision-making in empirical data rather than intuition or tradition.

4.3. Collaborative Approach and Pedagogical Innovation

It is important to emphasize that the methodology of this study reflects the principles of the "Students as Partners" approach, a key pedagogical trend in modern higher education (51). By basing the initial construction of the instrument on structured interviews with the students themselves, it was ensured that their voices and lived experiences formed the essential basis of the instrument. This collaborative process transcends the traditional model where students are passive research subjects, positioning them as active partners in improving their own educational environment—a practice that increases engagement and effectiveness in curriculum reforms.

4.4 Limitations of the Study

The study has limitations that must be considered. First, the use of non-probability sampling in the exploratory phases restricts the generalizability of findings to the entire student population of the Faculty of Medicine and Biology at the Autonomous University of Chihuahua (UACH). However, this methodological choice was deliberate: the main objective was not statistical generalization but contextual validity, ensuring that the language and themes resonated with the lived reality of students in the specific program. Second, the single-institution validation limits its direct applicability to other faculties with different institutional cultures, demographic profiles, or curricula, requiring additional validation studies. Nevertheless, the methodology used can serve as a reference for developing contextualized instruments. Finally, the qualitative nature of the instrument is both a strength and a limitation. Its open-ended questions generate detailed data ideal for comprehensive diagnosis, but they are not suitable for mass administration or complex statistical analysis. The instrument is excellent for answering the "what" and "how" of student problems, without allowing for the determination of quantitative associations between factors and numerical scores.

5. Conclusions

- A 38-item qualitative instrument was designed and rigorously validated for the systematic identification of intrinsic and extrinsic factors affecting the academic performance of medical students. Excellent content validity was achieved, resulting from a meticulous mixed-methods validation process that included expert review and pilot testing with the target population.
- The fundamental contribution of this research is to provide a holistic, contextualized, and validated diagnostic tool, addressing a gap identified both locally and within the broader landscape of medical education in Latin America, where the scarcity of instruments of this nature has been widely documented. Beyond its scientific value, this instrument represents a comprehensive institutional diagnostic tool with multiple applications that transcend traditional evaluation. Its systematic application generates an information ecosystem that simultaneously benefits students, faculty, and the institution as a whole, creating a virtuous cycle of continuous improvement based on empirical evidence.
- The ultimate value of this instrument lies in its ability to empower the institution, enabling it to move from conjecture to evidence-based action. It constitutes an essential step toward a deeper and more nuanced understanding of the student experience. In this way, it lays the groundwork for the future development of effective and targeted intervention strategies aimed at mitigating the risk of failing, reducing dropout rates, and ultimately promoting the success and overall well-being of the next generations of healthcare professionals.

Funding: No funding was available.

Declaration of conflict of interest: The authors declare that they have no conflict of interest.

Authors' contributions: Gustavo Domínguez Celis: Design, development, and writing of the article. Lorena Mendoza Anaya: Advising on the topic, design, and writing of the article. Review and editing of the article. Sandra Alicia Reza López: Design and statistical analysis of the study. Review and editing of the article. María Cecilia Ishida Gutiérrez: Review and editing of the article. Omar Fierro Fierro: Review and editing of the article.

6. References

1. Rubiano Romero SS, Martínez Huertas JC. Academic performance as a behavior in the teaching-learning process. *Ciencia Latina.* **2024,** 8(2), 5247-61. https://doi.org/10.37811/cl_rcm.v8i2.10941
2. Navarro RE. Academic performance: concept, research and development. *REICE.* **2003,** 1(2), 1-15. https://doi.org/10.15366/reice2003.1.2.007.

3. Nazir S, Khalid A, Yousaf D, et al. Potential dropout thoughts and their influencing factors among medical students. *Cureus.* **2024,** 16(11), e74757. https://doi.org/10.7759/cureus.74757.
4. Fernández MA, Gutiérrez D, Cruz P, et al. School dropout in medicine: analysis of three graduating classes. *FEM.* 2020, 23(6), 331-333. https://dx.doi.org/10.33588/fem.236.1094.
5. Martínez González A, Herrera Penilla CJ, García Minjares M, et al. Academic trajectories of undergraduate students in health sciences. *Gac Med Mex.* **2023,** 159(3), 238-244. http://dx.doi.org/10.24875/GMM.22000393.
6. Bautista Rodríguez G, Fortoul T. Academic trajectories of three generations of a Bachelor of Medicine during the COVID-19 pandemic. *RIEM.* **2025,** 14(53), 25-34. https://doi.org/10.22201/fm.20075057e.2025.53.24609.
7. Heffington DV, Dzay Chulim F, Fernández de Lara HR, et al. A Quantitative Approach to School Dropout in Higher Education in Southeastern Mexico. *Revista de la educación superior.* **2024,** 53(211), 19-40. https://doi.org/10.36857/resu.2024.211.2954.
8. Bin Abdulrahman KA, Khalaf AM, Bin Abbas FB, et al. Study habits of highly effective medical students. *Adv Med Educ Pract.* **2021,** 12, 627–33. https://doi.org/10.2147/AMEP.S309535.
9. Bin Abdulrahman KA, Alshehri AS, Alkhalifah KM, et al. The relationship between motivation and academic performance among medical students in Riyadh. *Cureus.* **2023,** 15(10), e46815. https://doi.org/10.7759/cureus.46815.
10. Guerrero López JB, Monterrosas AM, Reyes Carmona C, et al. Factors related to academic performance in medical students. *Health Ment.* **2023,** 46(4), 193–200. https://doi.org/10.17711/SM.0185-3325.2023.024.
11. Gruppen LD, Irby DM, Durning SJ, et al. Conceptualizing learning environments in the health professions. *Acad Med.* **2019,** 94(7), 969–74. https://doi.org/10.1097/acm.0000000000002702.
12. Jaber MH, Dafallah IA, Mohammed AY, et al. Socioeconomic disparities and their effect on medical student academic attainment Sudanese Universities. *BMC Med Educ.* **2024,** 24(1), 929. https://doi.org/10.1186/s12909-024-05867-4.
13. Rincon B, Bravo DY, Arnold E, et al. Community and family relationships across the transition to medical school: links to student adjustment. *Front Psychol.* **2024,** 15, 1330455. https://doi.org/10.3389/fpsyg.2024.1330455.
14. Schiekirka Schwake S, Anders S, Von Steinbüchel N, et al. Facilitators of high-quality teaching in medical school: findings from a nation-wide survey among clinical teachers. *BMC Med Educ.* **2017,** 17(1), 178. https://doi.org/10.1186/s12909-017-1000-6.
15. Rashid A, Yasmeen R, Ahmed R, et al. Factors leading to the academic failure of undergraduate medical students - Predict early to prevent. *Pak J Med Sci.* **2022,** 38(8), 2071-75. https://doi.org/10.12669/pjms.38.8.5951.
16. Ratnapalan S, Jarvis A. How to identify medical students at risk of academic failure and help them succeed? An interview with a medical educator. *MedSciEduc.* **2020,** 30, 989–94. https://doi.org/10.1007/s40670-020-00940-1.
17. Aguilar Bustamante JS, Cadena Durán MD, Jara Guerrero ER. Prevalence and impact of academic stress on medical students at the Technical University of Machala, 2024-DI. *Ciencia Latina.* **2024,** 8(6), 9815–29. https://doi.org/10.37811/cl_rcm.v8i6.15645.
18. Kilic R, Nasello JA, Melchior V, et al. Academic burnout among medical students: respective importance of risk and protective factors. *PublicHealth.* **2021,** 198, 187–95. https://doi.org/10.1016/j.puhe.2021.07.025.
19. Ilić IM, Ilić MD. The relationship between the burnout syndrome and academic success of medical students: a cross-sectional study. *Arh Hig Rada Toksikol.* **2023,** 74(2), 134–41. https://doi.org/10.2478/aiht-2023-74-3719.

20. Lavados Toro NA, Inzunza Carrasco S, Lillo Miranda B, et al. Evaluation of interventions to mitigate academic stress in medical students: a systematic review. *Rev Esp Edu Med.* **2025,** 6(1), 633961. https://doi.org/10.6018/edumed.633961.
21. Mirabal Martínez G, Román Rodríguez A, Silva Lago R, et al. Characterization of academic stress in medical students in the municipality of Bahía Honda. *Gac méd estud.* **2025,** 6, e493. https://revgacetaestudiantil.sld.cu/index.php/gme/article/view/493.
22. Salazar Flórez E, Arias Castro CE, Quintero Pinzón D, et al. Mental health in medical students: a challenge beyond academic stress. *Psicol. caribe.* **2024,** 41(1), 83–103. https://doi.org/10.14482/psdc.41.1.919.265.
23. Venkteshwar A, Maney K. The role of mental health in the medical performance of MBBS students. *Migrat. Lett.* **2024,** 21(S2), 638–47. https://migrationletters.com/index.php/ml/article/view/6685.
24. Reyes Seáñez MA, Ibáñez Bernal C, De La Rosa Ríos JE. Exploration of effective study strategies of competent students in Human Anatomy. *Exploratoris: Journal of Global Reality.* **2019,** 8(1), 8-14. https://www.academia.edu/44967588/Exploraci%C3%B3n_sobre_las_estrategias_de_estudio_efectivo_de_estudiantes_competentes_en_Anatom%C3%ADa_Humana_2019
25. Bracho Pernalete LC. Qualitative and quantitative validation of an instrument to measure university student satisfaction. *Educational Innovations.* **2021,** 23(35), 55–72. https://doi.org/10.22458/ie.v23i35.3590.
26. Homberg A, Rotter G, Thye M, et al. Instruments for evaluating undergraduate medical education in complementary and integrative medicine: a systematic review. *J Integr Complement Med.* **2025,** 31(7), 596-610. https://doi.org/10.1089/jicm.2024.0614.
27. Jensen Doss A, Hawley KM. Understanding barriers to evidence-based assessment: clinical attitudes toward standardized assessment tools. *J Clin Child Adolesc Psychol.* **2010,** 39(6), 885–96. https://doi.org/10.1080/15374416.2010.517169.
28. Barteit S, Guzek D, Jahn A, et al. Evaluation of e-learning for medical education in low- and middle-income countries: a systematic review. *Comput Educ.* **2020,** 145, 103726. https://doi.org/10.1016/j.compedu.2019.103726.
29. Fajardo Dolci GE, Santacruz Varela J, Lara Padilla E, et al. General characteristics of medical education in Mexico. A view from medical schools. *Salud Publica Mex.* **2019,** 61(5), 648-56. https://doi.org/10.21149/10149.
30. León Bórquez R, Lara Vélez VM, Abreu Hernández LF. Medical education in Mexico. *FEM.* **2018,** 21(3), 119-28. https://dx.doi.org/10.33588/fem.213.949.
31. Santander P. Why and how to do discourse analysis. *Cinta Moebio.* **2011,** 41, 207–24. http://dx.doi.org/10.4067/S0717-554X2011000200006.
32. Van Dijk TA. Discourse as Structure and Process. 3rd ed. Barcelona: *Gedisa Mexicana.* **2017,** 512 p. https://dialnet.unirioja.es/servlet/libro?codigo=10701
33. Urra E, Muñoz A, Peña J. Discourse analysis as a methodological perspective for health researchers. *Enferm Univ.* **2013,** 10(2), 50–57. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S166570632013000200004&lng=es.
34. Salomón RV. Sample size delimitation in pilot studies. *Rev Latinoam Metodol Investig Soc.* **2025,** 15(29), 7–21. https://relmis.com.ar/ojs/index.php/relmis/article/view/delmitacion_muestral.
35. García García JA, Reding Bernal A, López Alvarenga JC. Sample size calculation in medical education research. *Investig educ médica.* **2013,** 2(8), 217–24. http://dx.doi.org/10.1016/s2007-5057(13)72715-7.
36. Vera Ponce VJ, Zuzunaga Montoya FE, Huaman Vega CH, et al. Sample size calculation and precision for epidemiological studies: development and implementation of the Prevalence Calculator package in R. *Rev Peru Cienc Salud.* **2025,** 7(2), 156–63. http://dx.doi.org/10.37711/rpcs.2025.7.2.4.

37. Mayorga Ponce RB, Virgen Quiroz AK, Martinez Alamilla A, et al. Pilot Test. *ICSA.* **2020,** 9(17), 69-70. https://doi.org/10.29057/icsa.v9i17.6547.
38. Bujang MA, Omar ED, Foo DHP, et al. Sample size determination for conducting a pilot study to assess reliability of a questionnaire. *Restor Dent Endod.* **2024,** 49(1), e3. https://doi.org/10.5395/rde.2024.49.e3.
39. Whitehead AL, Julious SA, Cooper CL, et al. Estimating the sample size for a pilot randomized trial to minimize the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Stat Methods Med Res.* **2016,** 25(3), 1057–73. http://dx.doi.org/10.1177/0962280215588241.
40. Yusoff MSB. ABC of content validation and content validity index calculation. *Educ Med* J. **2019,** 11(2), 49–54. http://dx.doi.org/10.21315/eimj2019.11.2.6.
41. Galicia Alarcón LA, Balderrama Trápaga JA, Edel Navarro R. Content validity by expert judgment: Proposal for a virtual tool. *Apert.* **2017,** 9(2), 42–53. http://dx.doi.org/10.32870/ap.v9n2.993.
42. Vázquez Cancela O, Lens Perol G, Mascareñas Garcia M, et al. Assessing the Reliability and Validity of a Questionnaire Evaluating Medical Students' Attitudes, Knowledge, and Perceptions of Antibiotic Education and Antimicrobial Resistance in University Training. *Antibiotics.* **2024,** 13(12), 1126. https://doi.org/10.3390/antibiotics13121126.
43. Saenz Acuña AO, Parra Acosta H. Validation of an instrument that measures disruptive education in clinical training. *Rev Med Inst Mex Seguro Soc.* **2023,** 61(3), 274-82. https://revistamedica.imss.gob.mx/index.php/revista_medica/article/view/4776.
44. Herrera Masó JR, Calero Ricardo JL, González Rangel MA, et al. The expert consultation method in three levels of validation. *Rev haban cienc méd.* **2022,** 21(1), e4711. http://www.revhabanera.sld.cu/index.php/rhab/article/view/4711.
45. Escobar Pérez J, Cuervo Martínez Á. Content validity and expert judgment: an approach to its use. *Advances in measurement.* **2008,** 6, 27-36. http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf.
46. Carrera Farran FX, Vaquero Tió E, Balsells Bailón MÁ. Digital skills assessment instrument for adolescents at social risk. *EDUTEC.* **2011,** 35, a154. https://doi.org/10.21556/edutec.2011.35.410.
47. Hernández Nieto R. Data collection instruments in social sciences and biomedical sciences. 1st ed. Mérida, Venezuela: *Universidad de los Andes.* **2012,** 370 p. https://www.academia.edu/37886946/Instrumentos_de_recoleccion_de_datos_en_ciencias_sociales_y_ciencias_biomedicas_Rafael_Hernandez_Nieto_pdf
48. Maldonado Suárez N, Santoyo Telles F. Content validity by expert judgment: quantitative and qualitative integration in the construction of measurement instruments. *REIRE.* **2024,** 17(2), 1–19. https://doi.org/10.1344/reire.46238.
49. Roebianto A, Savitri SI, Aulia I, et al. Content validity: Definition and procedure of content validation in psychological research. *TPM.* **2023,** 30(1), 5-18. https://doi.org/10.4473/TPM30.1.1.
50. Martínez González A, García Minjares M, Zapata Castilleja CA, et al. Accreditation of medical education programs: comparison of results between evaluation teams. *Inv Ed Med.* **2024,** 13(49), 65–75. https://doi.org/10.22201/fm.20075057e.2024.49.23531.
51. Cook Sather A, Bovill C, Felten P. Engaging students as partners in learning and teaching: a guide for faculty. Nashville, TN: *John Wiley & Sons.* **2014,** 304 p. https://www.wiley.com/en-gb/Engaging+Students+as+Partners+in+Learning+and+Teaching%3A+A+Guide+for+Faculty-p-9781118434581