U.PORTO

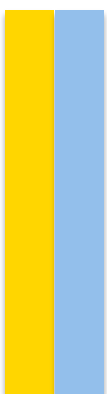
MESTRADO INTEGRADO EM MEDICINA

# Medical Education at the School of Medicine and Biomedical Sciences of the University of Porto: Student Trends and Perspectives

Pedro Dias Moreira

M

2025



# Medical Education at the School of Medicine and Biomedical Sciences of the University of Porto: Student Trends and Perspectives

Dissertação de candidatura ao grau de Mestre em Medicina, submetida ao Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto

Estudante:
Pedro Dias Moreira
Estudante do 6º ano do Mestrado Integrado em Medicina no Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto
Endereço eletrónico: 25pedromoreira09@gmail.com/up201906273@up.pt

Orientador: Milton Severo [a,c,d]

Coorientadora: Mariana P. Monteiro [b,d]

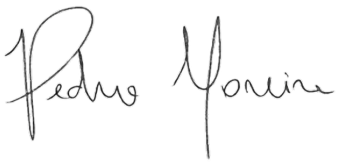
[a] Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto. (Rua de Jorge Viterbo Ferreira, 228, 4050-313, Porto, Portugal)

[b] Unit for Multidisciplinary Research in Biomedicine (UMIB), School of Medicine and Biomedical Sciences (ICBAS), University of Porto, Porto, Portugal. (Rua de Jorge Viterbo Ferreira n.º 228. 4050-313, Porto, Portugal)

[c] EPIUnit–Instituto de Saúde Pública, Universidade do Porto. (Rua das Taipas, 135, 4050-600 Porto, Portugal)

[d] Laboratório para a Investigação Integrativa e Translacional em Saúde Populacional (ITR). (Rua das Taipas, 135, 4050-600 Porto, Portugal)

O autor,



Pedro Dias Moreira

O orientador,

Assinado por: **Milton Severo Barros da Silva**
Num. de Identificação: 11469501
Data: 2025.06.05 15:22:30+01'00'

Professor Doutor Milton Severo

A coorientadora,

Professora Doutora Mariana Monteiro

Porto, junho de 2025

## Acknowledgements

Ao Professor Doutor Milton Severo, agradeço o acompanhamento rigoroso e o apoio essencial na definição metodológica e na análise estatística deste trabalho, bem como a clareza e objetividade com que orientou a interpretação dos resultados.

À Professora Doutora Mariana Monteiro, agradeço os contributos decisivos para a reflexão crítica e elaboração da discussão, assim como o acompanhamento atento e pertinente ao longo de todo o processo.

A ambos, o meu sincero reconhecimento pela orientação competente e dedicada.

## Resumo

**Introdução:** Os Inquéritos Pedagógicos da Universidade do Porto (IPUP) foram implementados em 2004 com o objetivo de monitorizar e melhorar a qualidade do ensino. Apesar dos avanços e das diversas revisões realizadas, a utilização destes instrumentos para análises longitudinais e, particularmente, em domínios específicos como a educação médica, continua pouco explorada. Este estudo procura colmatar essa lacuna através de uma análise longitudinal da opinião dos estudantes de medicina ao longo de vários anos letivos consecutivos.

**Objetivo:** Caracterizar as tendências e padrões de avaliação dos estudantes sobre o plano curricular do Mestrado Integrado em Medicina do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto (ICBAS-UP), analisando as respostas aos IPUP recolhidas ao longo de três anos letivos consecutivos (2020 a 2023).

**Métodos:** O IPUP avalia nove subdomínios do ensino e aprendizagem: apreciação e clareza, dificuldade, apoio à autonomia, consistência e ajuda, estrutura, relacionamento, efeitos da unidade curricular (UC), envolvimento e avaliação. Foram analisadas as respostas agregadas de 74 unidades curriculares ao longo de três anos (total de 221 observações). A taxa de resposta global aos IPUP variou entre 42,55% e 49,77% no 1.º semestre e entre 35,18% e 36,15% no 2.º semestre, ao longo dos três anos analisados. Para a análise das tendências ao longo do tempo, recorreu-se à utilização de modelos lineares mistos.

**Resultados:** As avaliações dos estudantes foram globalmente positivas em todos os subdomínios (médias superiores a 5 numa escala de 1 a 7). Verificaram-se diferenças estatisticamente significativas entre os anos letivos, sendo o 3.º ano o que registou as classificações mais baixas (média=5,50±0,23; dificuldade=5,41±0,75) e o 4.º ano o melhor classificado (média=5,94±0,15; dificuldade=5,12±0,85). Por área científica, a Patologia foi a melhor avaliada (média=6,38±0,37; dificuldade=4,53±0,26), enquanto a Semiologia (média=5,27±0,48; dificuldade=5,78±0,12) e a Farmacologia (média=5,28±0,56; dificuldade=6,45±0,21) apresentaram os resultados mais baixos. Observou-se uma correlação positiva entre a média das classificações finais e as avaliações dos subdomínios, com exceção da dificuldade, que apresentou uma correlação negativa (r = -0,57). Um maior número de créditos ECTS (*European Credit Transfer and Accumulation System*) revelou também uma associação positiva com as perceções dos estudantes. Ao longo do período avaliado, verificou-se uma diminuição significativa nas avaliações do subdomínio Docente (β = -0,147, IC=95%), enquanto

os domínios Unidade Curricular e Dificuldade se mantiveram estáveis. O Coeficiente de Correlação Intraclasse (ICC) variou entre 0,45 (Docente) e 0,86 (Dificuldade), indicando um elevado nível de consistência nas perceções dos estudantes.

**Conclusão:** Este estudo fornece evidência sobre o potencial dos IPUP como ferramenta de avaliação longitudinal da perceção dos estudantes relativamente à qualidade da formação médica pré-graduada. A análise permitiu identificar pontos fortes, fragilidades e oportunidades de melhoria na formação médica, segundo a perspetiva discente. As avaliações dos estudantes foram globalmente positivas em todos os subdomínios, embora haja diferenças significativas entre anos letivos e domínios científicos. Apesar destes resultados exigirem uma análise mais aprofundada, incluindo a visão dos docentes, a consistência dos dados ao longo dos anos sugere, por um lado, forte fiabilidade dos mesmos e, por outro, que as oportunidades de melhoria não foram aproveitadas nem traduzidas em alterações significativas, revelando um potencial subaproveitado destes instrumentos para a melhoria contínua do ensino e aprendizagem. A deterioração progressiva da perceção do Domínio Docente ao longo do tempo constitui um sinal de alerta, que poderá representar uma ameaça à qualidade do ensino, caso não seja adequadamente monitorizada e invertida.

**Palavras-chave:** educação médica, pré-graduada, currículo, inquéritos pedagógicos, ambiente de aprendizagem, avaliação dos estudantes

## Abstract

**Introduction:** The Pedagogic Surveys (IPUP) were introduced in 2004 at the University of Porto to monitor and improve teaching quality. Despite the advancements and implementation of several revisions, the use of these tools for longitudinal analysis, and particularly in specific knowledge domains, such as medical education, remains underexplored. This study seeks to address this gap by conducting a longitudinal analysis of medical students' feedback over consecutive academic years.

**Objective:** To characterize the students' evaluation trends and patterns of undergraduate curriculum of the Integrated Master's in Medicine of the School of Medicine and Biomedical Sciences at the University of Porto (ICBAS, UP) by analysing the IPUP responses collected along three consecutive academic years spanning from 2020 until 2023.

**Methods:** The IPUP are designed to evaluate nine subdomains of teaching and learning, namely: appreciation and clarity, difficulty, support for autonomy, consistency and help, structure, relationship, effects of the UC, engagement, and assessment. The study analysed clustered responses to IPUP to 74 Curricular Units (UC) over three years, rendering 221 observations (74 UCx3years). The overall IPUP response rate ranged from 49.77% to 42.55% in the first semester and from 36.15% to 35.18% in the second semester over the three academic years analysed. Linear mixed-effects models were employed to assess temporal trends.

**Results:** Student evaluation scores were overall positive across all subdomains (mean scores > 5 on a 7-point scale). Statistically significant differences were found between academic years, with the 3rd year consistently presenting the lowest scores (mean=5.50±0.23 and difficulty=5.41±0.75) and the 4th year showing the highest (mean=5.94±0.15 and difficulty=5.12±0.85). When analysed by scientific domain, Pathology was the one with the highest scores (mean=6.38±0.37 and difficulty=4.53±0.26), while Semiology (mean=5.27±0.48 and difficulty=5.78±0.12) and Pharmacology (mean=5.28±0.56 and difficulty=6.45±0.21) received the lowest scores. A significant positive correlation was observed between average final grade and subdomain evaluations, except for difficulty, which showed a negative correlation ($r = -0.57$). A higher number of ECTS (European Credit Transfer and Accumulation System) credits was also positively associated with student perceptions. During the evaluation period, a significant decline was observed in the Teacher domain ($\beta = -0.147$, CI=95%), whereas the UC and Difficulty domains remained stable. Intraclass Correlation Coefficient (ICC) values

ranged from 0.45 (Teacher) to 0.86 (Difficulty), indicating a high level of consistency in student perceptions.

**Conclusion:** This study provides important insights into the potential use of the IPUP as a tool for longitudinal assessment of the student perspectives of undergraduate medical education quality. This analysis allowed us to pinpoint and highlight the strengths, weaknesses, and opportunities for improvement in medical education according to the students' perspectives. Student evaluation scores were overall positive across all subdomains, although there were significant differences between academic years and scientific domains. Although these findings require a more comprehensive analysis, including considering faculty perspectives, the consistency of results over the years suggests, on one hand, strong reliability and, on the other, that the opportunities for improvement were not taken into account or led to the implementation of significant changes, denoting the missed opportunity provided by these tools for the refinement of teaching and learning processes. Of particular note is the progressive decline in student perceptions of the Teacher Domain over time, which can be a potential threat to educational quality if not promptly addressed.

**Keywords:** medical education, undergraduate, curriculum, pedagogical surveys, learning environment, student evaluation

## List of Abbreviations

ANOVA: Analysis of Variance

β: Regression Coefficient

CI: Confidence Interval

ECTS: European Credit Transfer and Accumulation System

ICBAS: School of Medicine and Biomedical Sciences of the University of Porto

ICC: Intraclass Correlation Coefficient

IPUP: Pedagogical Survey of the University of Porto

M (SD): Mean (Standard Deviation)

PCA: Principal Component Analysis

$R^2$: Coefficient of Determination

SC: Study Cycle

SIGARRA: Sistema de Informação para Gestão Agregada dos Recursos e dos Registos Académicos (University of Porto Academic Management Information System)

UC: Curricular Unit

UP: University of Porto

# Index

## List of Tables

## List of Figures

## Introduction

Obtaining pedagogical information about the Unit Course (UC) of Study Cycles (SC) is essential for ensuring the overall quality of education, contributing to the continuous improvement of educational processes. This information is typically gathered through questionnaires, known as Pedagogical Surveys, which enable students to critically and comprehensively evaluate various pedagogical aspects, ranging from the content of curricular programs to the teaching performance within the UC (1).

The most frequently assessed teaching competencies include individual pedagogical practices, the organization of the teaching-learning process, and student-teacher interaction (2). Among the key aspects analysed in a UC are its overall impact, the level of student engagement, the clarity and coherence of content and methodologies used, and the evaluation methods adopted (3).

Understanding how students perceive their learning environment is particularly relevant during key transitional periods, such as the shift from preclinical to clinical training, where perceptions worsen, especially in the third year, with partial recovery in the fourth year (4). Over the past decades, student feedback has emerged as an essential tool for assessing the learning environment (5). The Pedagogical Surveys of the University of Porto (IPUP) have been developed as part of an ongoing effort to improve educational quality. First introduced in 2004, the IPUP underwent several revisions, culminating in its latest version for the academic year 2016/2017 (6).

The results, available on the SIGARRA platform, have broad application potential: supporting evidence-based pedagogical decision-making, helping recognize teaching performance, and informing the analysis of SCs by relevant governing bodies, such as Pedagogical Councils and Organic Unit Boards (7). Additionally, the data helps to identify persistent weaknesses and promote continuous improvements, contributing to the strategic development of each Organic Unit and, consequently, the University.

Despite significant advancements in developing tools to gather pedagogical information from students, several gaps remain. The large volume of data collected through the IPUP presents a major challenge in terms of integration and analysis from a holistic perspective. While the survey results are presented quantitatively, with automatic statistical analysis providing an overview of teaching performance and curricular content, this approach alone is insufficient to translate students' perceptions into specific actions for improvement (8). The lack of academic literature on the effectiveness of IPUP, particularly in longitudinal contexts or specific scientific

domains such as medical education, hinders the development of robust methodologies for evaluating the actual impact of these tools on teaching quality (9). This academic gap highlights the necessity for further research to deepen the understanding of the relationship between student feedback and subsequent pedagogical developments, while also providing clear directions for enhancing educational practices based on this feedback.

Furthermore, the scarcity of longitudinal studies limits our ability to fully understand how student trends and perspectives evolve, and how these changes influence faculty development.

In light of these issues, the objective of this study was to conduct an integrated and longitudinal analysis over three years of the responses to the IPUP collected from students of the Integrated Master's in Medicine at the School of Medicine and Biomedical Sciences of the University of Porto (ICBAS), during the academic years 2020-2021, 2021-2022, and 2022-2023. The main aim was to identify trends, patterns, and opportunities for improvement related to the program's curriculum.

## Methods

In the academic years 2020-2021, 2021-2022, and 2022-2023, students from the Integrated Master's in Medicine at ICBAS were invited to respond to the IPUP to evaluate the UC attended. These years were selected for analysis because they follow the implementation of a major curricular reform, which was fully introduced in 2019–2020. However, that initial year was excluded due to the atypical academic conditions imposed by the COVID-19 pandemic, which may have significantly influenced students' perceptions and responses. Participation in the surveys was voluntary and anonymous, with the questionnaires being made available online through the SIGARRA platform. The surveys were conducted every semester for all UCs with more than 10 enrolled students, ensuring a minimum level of response representativeness (6). IPUP responses were collected across 74 UCs, yielding 221 observations in total. The overall IPUP response rates for the academic years 2020/2021, 2021/2022 and 2022/2023 were 49.77%, 44.93% and 42.55% in the first semester and 35.91%, 36.15% and 35.18% in the second semester, respectively.

The IPUPs are designed as an essential tool for monitoring and continuously improving teaching quality. These were first implemented in 2004 and undergo regular revisions, with the latest version dating from 2016/2017. Its development was based on pedagogical principles aimed at systematically assessing key aspects of students' academic experience. (6)

Although the survey includes an open-response field aimed at providing deeper insights into the strengths and weaknesses of the UC, allowing students to contextualize their scores and identify aspects not captured through numerical responses, this study only considered the responses to the closed-ended questions. Open-ended responses were excluded from the analysis to maintain consistency and focus on quantitative evaluation.

The closed-ended questions used in this study consist of 12 questions organized into 9 subdomains, each with a specific objective: appreciation and clarity, which assesses the organization and coherence of the information conveyed by lecturers, as well as their communication skills; difficulty, which measures students' perception of the UC's level of demand, considering factors such as workload and content complexity; support for autonomy, which analyses the extent to which the UC promotes independent study by providing materials and resources that encourage autonomous learning; consistency and support, which examines the alignment between the taught content, the accessibility of lecturers for clarifying doubts, and the support provided throughout the semester; structure, which evaluates the overall organization of the UC, including the coherence of syllabi, class distribution, and content progression throughout the semester; relationship, which examines student-lecturer interaction, considering the availability of faculty member, pedagogical dynamics, and the creation of a positive learning environment; effects of the Unit Course, which identifies the UC's impact on students' academic and professional development, analysing its relevance in medical training; engagement, which measures the level of student participation and motivation in UC activities, including their interest in the topics covered; and assessment, which analyses the clarity and fairness of grading criteria, as well as the adequacy of assessment methods to the taught content (6).

Each question is answered on a Likert scale from 1 (very low) to 7 (very high), allowing for a detailed quantitative analysis of UC and lecturer performance. The questions that comprise the surveys, including their structure and content details, can be found in a previous publication (6).

## Statistical Analysis

The subdomains were described and compared by curricular year and scientific area with mean and standard deviation (SD) and using Analysis of Variance (ANOVA) or Kruskal-Wallis tests. The Pearson correlation was used to assess the magnitude of the linear association between the subdomains and the average grade of the students.

To derive summary domains that could be used to explain the 9 subdomains, a principal components analysis (PCA) with oblimin rotation was performed. Items with factor loadings with

a magnitude of at least |0.40| were in specific component. The number of dimensions for each scale was estimated using optimal coordinates and Kaiser's criterion (eigenvalues > 1). Standardized factor scores produced by the PCA were used in subsequent analyses (Table I).

Linear mixed effects models were used to estimate the linear time-trend of domains scores Unit Course, Teacher and Difficulty.

The first model was built only with the year of assessment as a fixed effect and random effects for UC. The second model added the average grade of the students of UC as a fixed effect. The third model additionally included the number of ECTS and the curricular year. To compare the fit between models, the log likelihood ratio test, the Bayesian information criterion, and the Akaike information criterion were used. The fixed effects were expressed as linear regression coefficients and corresponding 95% confidence intervals (CI). The random effects (measures of variation) were expressed as Intraclass Correlation Coefficient (ICC). The ICC provides an estimate of the proportion of total variance related to UC.

A sensitivity analysis was performed with the same parametrization of the models using the subdomains (Appendix).

All analyses were done using the software R. 2023.09.0+463 version and the following libraries were used: ggplot2, ggpubr, and nlme.

## Results

### Description of Subdomains According to Different Characteristics of the UC

#### General Overview of the Evaluations

The scores attributed to all subdomains were overall positive (Table II), considering that the scale ranges from 1 to 7. The average values across subdomains were consistently above 5.2, indicating a positive perception of the evaluated dimensions. The lowest mean scores were attributed to the effects of the UC subdomain in the 1st year (mean = 5.20), while the highest scores were attributed to the engagement subdomain in the 6th year (mean = 6.41).

#### Variation in Perceptions Across Academic Years

Moreover, Table II shows that there are statistically significant differences in the various subdomains across the different academic years, except for difficulty (P = 0.33). The trend reveals that the highest scores across most subdomains are attributed to the last three years of the course, particularly the 4th year, which consistently presents the highest scores across several subdomains. Conversely, the first three years show lower scores, with the 3rd year presenting the lowest scores across most subdomains.

**Differences Between Scientific Domains**

Although the scores attributed to all subdomains were overall positive, considering that the scale ranges from 1 to 7, statistically significant differences were still found across the various subdomains among scientific areas ($P < 0.001$) (Table III). The highest scores were attributed to Pathology, which consistently presents the highest mean values for most subdomains, such as support for autonomy, consistency and support, structure, relationship, difficulty, and assessment. Conversely, Semiology and Pharmacology received consistently lower scores in several subdomains, and most particularly, assessment and difficulty.

**Association with Final Grade Average**

The evaluated subdomains showed a positive correlation ($P < 0.05$) with the final grade of each Unit Course, as shown in Figure 1. The graphs show that, except for difficulty, where the linear trend exhibits a negative correlation ($r = -0.57$), indicating that higher final grades are associated with lower perceptions of difficulty, all other subdomains display positive correlations. This suggests that higher final grades are related to higher IPUP scores, as well.

**Association with the Number of ECTS**

There are differences between the subdomains and the number of ECTS for the subdomains: appreciation and clarity, difficulty, effects of the UC, and engagement ($P < 0.05$), as depicted in Figure 2. In all cases where the association was statistically significant, a positive relationship was observed, meaning that the average scores attributed to these subdomains tend to increase with the number of ECTS. This suggests that UCs with more ECTS are associated with better perceptions in the above-mentioned subdomains.

## Trends of Summary Domains

**Longitudinal Trends: Linear Mixed Models**

A positive association between the UC and Teacher domains was identified, additionally, regardless of the academic year, the relative position of the Unit Courses remained stable (Figure 3). This pattern suggests consistency of scores attributed over time. The intraclass correlation coefficients (ICC) ranged from 0.51 for the Teacher domain to 0.86 for the Difficulty domain, demonstrating a moderate to high degree of consistency in the scoring of these domains.

**Teacher Domain**

As shown in Table IV, in the Teacher Domain, in terms of raw association, a significant decrease in scores was observed over time ($\beta = -0.162$; $P = 0.006$), indicating a decline in students' perception of this domain over time. In Model 1, adjusted for final grades, the

decrease in scores remained significant (β = -0.147; P = 0.011), suggesting that the reduction in perception of the Teacher Domain is not fully explained by the UC final scores. However, a significant positive association was found between final grades and the Teacher Domain (β = 0.209; P < 0.001), indicating that higher scores are consistently associated with more favourable perceptions of this dimension.

In Model 2, further adjusted for ECTS and academic year, the downward trend remained statistically significant (β = -0.147; P = 0.011), reinforcing that the decline in perception of the Teacher Domain is independent of these factors. On the other hand, ECTS showed no significant association (P = 0.871), suggesting that the credit load of UCs does not influence students' perceptions related to the Teacher Domain. Moreover, the coefficient for the 6th academic year (β = -0.935; P = 0.028) indicates a sharper decline in scores attributed to this year compared to the 1st year, suggesting a worsening of student perceptions along the academic years.

The ICC of Model 2, at 0.45, indicates that 45% of the variability in perceptions of the Teacher Domain can be attributed to differences between courses. These results highlight the consistency of declining students' perceptions of the Teacher Domain over time.

**Unit Course Domain**

According to Table V, in terms of gross association, in the UC Domain, no statistically significant change in the scores over time was observed (β = -0.024; P = 0.555), indicating stability in students' perception of this dimension. In Model 1, adjusted for the final grade, no significant association was found (β = -0.009; P = 0.813), suggesting that the stability in the perception of the UC Domain remains independent of the grade obtained by the student. However, a positive and significant association was observed between the final grades and the UC Domain (β = 0.218; P < 0.001), indicating that higher final grades are associated with more favourable student perceptions of this dimension. In Model 2, additionally adjusted for the number of ECTS and the curricular year, the association between time and the UC Domain remained non-significant (β = -0.007; P = 0.869), reinforcing the hypothesis that perception does not vary significantly over the academic years. On the other hand, the number of ECTS showed a positive and statistically significant association with the UC Domain (β = 0.437; P < 0.001), indicating that UCs with more credits gain more favourable perceptions from students. Regarding the curricular year, no significant associations were found in any of the years.

The ICC of Model 2, at 0.66, indicates that 66% of the variability in perceptions of the UC Domain can be attributed to differences between UC, suggesting a stable perception of the UC Domain over time.

**Difficulty Domain**

As evidenced by Table VI in terms of gross association, regarding the Difficulty Domain, no statistically significant change in score over time was observed (β = -0.003; P = 0.900 in Model 0), indicating that students' perception of difficulty remained stable over the academic years. In Model 1, adjusted for final grade, the association with time remained non-significant (β = -0.017; P = 0.394). However, a negative and statistically significant association was observed between final grade and perceived Difficulty (β = -0.211; P < 0.001), suggesting that higher final grades are associated with lower perceptions of difficulty in UC. In Model 2, additionally adjusted for the number of ECTS and curricular year, the association with time remained non-significant (β = -0.017; P = 0.392), reinforcing the stability of the perception of difficulty over time. On the other hand, the number of ECTS showed a positive and significant association with Difficulty (β = 0.256; P < 0.001), indicating that UCs with a higher credit load are perceived as more difficult. Across curricular years, only the 5th year showed a statistically significant association (β = 0.452; P = 0.013), suggesting a higher perception of difficulty in this specific year compared to the 1st year.

The ICC of Model 2, at 0.75, indicates that 75% of the variability in perceptions of Difficulty can be attributed to differences between UC, highlighting that the perception of Difficulty is relatively stable over time.

## Discussion

The main aim of this study was to conduct a longitudinal analysis of student perceptions regarding UCs in the Integrated Master's in Medicine at ICBAS, based on responses to the IPUP over three academic years.

The findings of this study indicate that student perceptions of UCs are overall good, although these vary across the Medical Course. The third year, coinciding with the transition from preclinical to clinical education, consistently presents as the most critical for students' perceptions, with the lowest scores attributed in the survey. This phenomenon is extensively documented in the literature, highlighting specific challenges of medical education during this transition phase, such as course structure, coherence of content, and the appropriateness of assessments (4, 10, 11). Third-year medical students often face a greater workload, a more fragmented and less interactive teaching environment, contributing to negative perceptions of the educational environment. Additionally, the pressure to rapidly acquire fundamental clinical competencies further exacerbates these negative perceptions (12). However, a significant improvement in students' perceptions occurs in the fourth year, marking the transition of

medical education to the clinical setting. This improvement is attributed to clearer learning objectives, a more practical teaching oriented toward clinical application, a greater sense of professional purpose, and more specific feedback received (4, 11, 13). Moreover, the improvement in students' perceptions can also be explained by a stronger sense of belonging and professional identity, facilitated by direct engagement with real-world clinical scenarios (10).

This study demonstrates a significant positive correlation between the average final grade of the UC and students' evaluations across most of the subdomains assessed, with the notable exception of difficulty, which shows a negative correlation that suggests that UCs with a greater workload and perceived difficulty to achieve higher grades, are associated with less positive perceptions (14). This pattern indicates that UCs with higher final grades receive more favourable student feedback. This phenomenon may result from the alignment between instructional content and assessment methods, prompting students to rate the courses more positively when they reach better academic achievements (4). However, there is a potential bias in student IPUPs scoring, where UCs with greater academic performance receive higher scores regardless of actual educational quality (1, 15). Such bias may distort perceptions of genuine educational quality.

The results suggest that UC with a higher number of ECTS are associated with more positive student perceptions, possibly because they correspond to subjects considered as core in the curriculum. Due to the greater workload, these UCs generally provide more time for in-depth learning, allowing for better consolidation of acquired knowledge, as well as offering practical activities and continuous feedback, all elements that are positively valued by students (16, 17).

Regarding the division by Scientific Domains, Pathology consistently received the highest scores, while Semiology and Pharmacology showed the lowest, particularly in assessment and difficulty. The favourable evaluations in Pathology can be attributed to effective teaching methods and successful integration between basic and clinical sciences. Such curricular alignment promotes a stronger understanding and a more positive learning experience, particularly when theoretical introductions are followed by practical microscope sessions and when normal and pathological tissue structures are taught in close association, as highly valued by students (18, 19). Conversely, the lower scores in Semiology and Pharmacology may reflect structural and pedagogical challenges. Semiology, for instance, faces difficulties in providing adequate clinical supervision and organising practical internships, with a negative impact on students' perceptions. High student-tutor ratios, such as those observed at ICBAS in the 3rd to 5th years (9.15 students per tutor), have been significantly associated with lower satisfaction in

clinical training (20, 21). Ratios above 6 students per tutor may compromise the quality of supervision, limit student engagement and reduce the overall effectiveness of clinical education (22). As for Pharmacology, the absence of interactive teaching methods and practical integration is likely to contribute to the lower scores attributed by students. Problem-Based Learning (PBL) strategies have been shown to improve student performance and perceptions by promoting more active and practical learning (23, 24). These findings suggest that the most positively evaluated Scientific Domains are those that provide structured, relevant teaching with effective methodologies that facilitate theoretical and practical integration. Interactive methods and adequate supervision emerge as critical aspects for enhancing student satisfaction and positive perceptions of the learning process (1, 16, 25).

The ICC, which estimates the proportion of total variance in scores attributable to differences between courses rather than random error or within-course variation, reveals considerable stability in scores over the years. This indicates appreciable consistency and reliability in the assessment process. Highly-rated UCs consistently receive positive evaluations, while lower-rated ones persist in being poorly evaluated. This stability can be interpreted in two ways: on one hand, it reflects solid coherence in students' perceptions, ensuring that the results are reliable and reflect sustained opinions. On the other hand, it suggests that pedagogical interventions have not produced measurable effects, possibly indicating resistance to change or inadequacy of the strategies implemented to enhance teaching quality (4).

The results also demonstrate a significant decline in students' perceptions of the Teacher Domain over time, even after adjusting for final grades, ECTS, and academic year. This finding suggests the existence of additional factors negatively influencing the perceived quality of teaching, factors that are not explained by these academic or curricular variables alone. Such a decline may reflect challenges related to teachers' resistance or difficulty in adapting to innovative and interactive teaching methodologies, increasingly valued by students (25). Furthermore, the lack of continuous and adequate training for faculty in modern pedagogical techniques that limit teaching effectiveness is another factor that can negatively impact student perceptions (1, 26). Perceived teaching quality tends to decline when teachers fail to meet students' rising expectations, particularly in the current educational context, where there is greater appreciation for continuous feedback, active learning methodologies and theoretical-practical integration (27). In contrast, the stability observed in the UC and Difficulty domains suggests a consistent pedagogical experience, perceived positively and aligned with student expectations, particularly when associated with higher final grades. Such consistency indicates that students perceive the curricular structure, academic demands and pedagogical approaches within these areas as appropriate and effective. These observations carry clear implications for

institutional strategy: there is an urgent need to evaluate the effectiveness of pedagogical interventions, especially those targeting teacher training and professional development. Continuous training programs, fostering a culture of regular and constructive feedback, and consistent adoption of active teaching methodologies, such as PBL and flipped classrooms, emerge as essential strategies to reverse the observed decline in students' perceptions of teaching quality (28, 29). In parallel, constraints that limit the institution capacity to recruit and retain talented faculty members may hinder the renewal and continuous improvement of teaching practices and must be adequately addressed as part of any strategy aimed at enhancing the quality of medical education (30).

This study presents some limitations. The three years analysed may be considered short, which may not be sufficient to assess significant changes. The quality of IPUP may compromise the reliability of the results, particularly due to the lack of questions specifically tailored to the Medical Course. Additionally, the study focuses exclusively on students' perceptions, disregarding the perspective of faculty members and objective evaluations. External factors may have influenced the results, particularly the previously mentioned performance-related bias, where UCs with higher final grades are evaluated more favourably. Finally, the exclusive use of quantitative data limits the understanding of the underlying reasons behind students' perceptions.

In summary, this study offers a novel perspective on the evolution of students' perceptions of medical education over three academic years, using an institutional evaluation tool that, despite being widely implemented, remains largely underexplored. The analysis revealed consistent trends in the evaluation of UCs, with the third year standing out as critical in the medical training pathway. Clear associations were identified between final grades and more favourable student perceptions. Additionally, a striking stability in patterns was observed across years and unit courses. These findings suggest that, despite the volume of feedback collected annually, its potential to guide pedagogical improvement has not been fully leveraged. The consistently lower scores attributed to the Teacher Domain pinpoint a critical need for faculty development and renewal. Therefore, this research contributes to a deeper understanding of the value, limitations and opportunities associated with the systematic use of pedagogical surveys in undergraduate medical education.

## Tables

Table I. Principal Component Analysis (PCA) Loadings and Communality Measures by Subdomain

TC1, TC2, TC3 – Factor loadings for the first, second and third latent traits, respectively.

h2 – Communality, representing the proportion of each variable's variance explained by the factors.

u2 – Uniqueness, representing the proportion of variance not explained by the factors.

com – Complexity index, indicating how much a variable is associated with multiple factors

| ## | Subdomain | TC1 | TC2 | TC3 | h2 | u2 | com |
|---|---|---|---|---|---|---|---|
| ## Consistency and support | 2 | 0.99 | | | 0.93 | 0.067 | 1.0 |
| ## Structure | 3 | 0.96 | | | 0.93 | 0.075 | 1.0 |
| ## Support for autonomy | 1 | 0.96 | | | 0.92 | 0.078 | 1.0 |
| ## Relationship | 4 | 0.95 | | | 0.93 | 0.068 | 1.0 |
| ## Effects of the Unit Course | 9 | | 0.96 | | 0.91 | 0.089 | 1.1 |
| ## Engagement | 5 | | 0.83 | 0.38 | 0.81 | 0.190 | 1.4 |
| ## Appreciation and clarity | 6 | 0.33 | 0.72 | | 0.85 | 0.151 | 1.4 |
| ## Assessment | 7 | 0.44 | 0.53 | | 0.79 | 0.212 | 2.4 |
| ## Difficulty | 8 | | | 0.97 | 0.93 | 0.069 | 1.0 |

Table II. Description of Subdomains According to Academic Year
Bold: Indicates the highest and lowest scores

| **Subdomain/Academic Year** | **1** | **2** | **3** | **4** | **5** | **6** | **P** |
|---|---|---|---|---|---|---|---|
| Support for autonomy, M (SD) | 5.75 (0.53) | 5.67 (0.57) | **5.65 (0.60)** | **6.06 (0.55)** | 5.96 (0.33) | 5.66 (0.88) | 0.001 |
| Consistency and support, M (SD) | 5.93 (0.44) | 5.82 (0.54) | **5.66 (0.60)** | **6.08 (0.56)** | 6.06 (0.43) | 5.72 (0.94) | 0.004 |
| Structure, M (SD) | 5.74 (0.48) | 5.68 (0.49) | 5.69 (0.54) | **6.10 (0.53)** | 5.93 (0.33) | **5.59 (0.84)** | <0.001 |
| Relationship, M (SD) | 5.75 (0.52) | 5.70 (0.64) | 5.60 (0.61) | 6.00 (0.65) | **6.01 (0.46)** | **5.55 (0.99)** | 0.003 |
| Engagement, M (SD) | **5.46 (0.62)** | 5.53 (0.49) | 5.68 (0.46) | 5.80 (0.60) | 5.93 (0.43) | **6.41 (0.48)** | <0.001 |
| Appreciation and clarity, M (SD) | 5.44 (0.84) | 5.61 (0.73) | **5.41 (0.53)** | **5.98 (0.44)** | 5.80 (0.43) | 5.55 (0.71) | <0.001 |
| Difficulty, M (SD) | 5.23 (0.78) | 5.21 (0.62) | 5.41 (0.75) | **5.12 (0.85)** | 5.34 (0.49) | **5.43 (0.61)** | 0.33 |
| Effects of the Unit Course, M (SD) | **5.20 (0.61)** | 5.35 (0.51) | 5.31 (0.51) | **5.71 (0.36)** | 5.54 (0.34) | 5.66 (0.72) | <0.001 |
| Assessment, M (SD) | 5.41 (0.85) | 5.54 (0.71) | **4.98 (0.89)** | **5.76 (0.69)** | 5.51 (0.46) | 5.02 (1.12) | <0.001 |

Table III. Description of Subdomains by Scientific Domains
Bold: Indicates the highest and lowest scores

| **Subdomain/ Scientific Area** | **ANATOMY** | **SURGERY** | **EPIDEMIOLOGY** | **PHARMACOLOGY** | **PHYSIOLOGY** | **GENETICS** | **MEDICINE** | **MICROBIOLOGY** | **PATHOLOGY** | **MENTAL HEALTH** | **SEMIOLOGY** | **P** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Support for autonomy, M (SD) | 5.87 (0.37) | 5.55 (0.65) | 5.92 (0.31) | 5.78 (0.21) | 5.83 (0.19) | 5.89 (0.78) | 6.01 (0.35) | 5.23 (0.42) | **6.58 (0.28)** | 5.68 (0.87) | **5.11 (0.31)** | <0.001 |
| Consistency and support, M (SD) | 5.95 (0.37) | 5.63 (0.70) | 5.78 (0.39) | 5.71 (0.20) | 6.00 (0.19) | 6.22 (0.51) | 5.97 (0.26) | 5.28 (0.39) | **6.70 (0.18)** | 5.70 (1.01) | **5.00 (0.33)** | <0.001 |
| Structure, M (SD) | 5.83 (0.38) | 5.58 (0.57) | 5.80 (0.31) | 5.81 (0.21) | 5.80 (0.17) | 5.97 (0.57) | 5.93 (0.29) | 5.24 (0.35) | **6.61 (0.26)** | 5.66 (0.82) | **5.12 (0.36)** | <0.001 |
| Relationship, M (SD) | 5.88 (0.37) | 5.57 (0.62) | 5.78 (0.43) | 5.54 (0.20) | 5.99 (0.27) | 6.04 (0.72) | 5.92 (0.30) | 5.24 (0.40) | **6.69 (0.21)** | 5.62 (0.97) | **5.11 (0.34)** | <0.001 |
| Engagement, M (SD) | 5.92 (0.29) | 6.07 (0.31) | 5.72 (0.42) | 5.54 (0.23) | 5.96 (0.15) | **5.36 (0.64)** | **6.52 (0.26)** | 5.53 (0.34) | 5.66 (0.31) | 5.88 (0.67) | 6.15 (0.20) | <0.001 |
| Appreciation and clarity, M (SD) | 6.12 (0.38) | 5.85 (0.30) | 5.06 (0.47) | 4.99 (0.28) | **6.34 (0.22)** | 5.43 (0.72) | 6.11 (0.25) | **4.93 (0.46)** | 6.33 (0.19) | 5.39 (0.87) | 5.62 (0.29) | <0.001 |
| Difficulty, M (SD) | 5.72 (0.25) | 5.49 (0.23) | 4.70 (0.45) | **6.45 (0.21)** | 5.58 (0.53) | 4.98 (0.72) | 6.31 (0.29) | 5.72 (0.31) | **4.53 (0.26)** | 4.68 (0.54) | 5.76 (0.12) | <0.001 |
| Effects of the Unit Course, M (SD) | 5.63 (0.32) | 5.57 (0.23) | 5.44 (0.45) | **4.70 (024)** | 5.83 (0.12) | 5.20 (0.71) | 5.83 (0.27) | 5.01 (0.34) | **5.87 (0.19)** | 5.56 (0.64) | 5.60 (0.24) | <0.001 |
| Assessment, M (SD) | 5.92 (0.45) | 5.33 (0.53) | 5.06 (0.81) | **4.15 (0.35)** | 6.02 (0.28) | 5.63 (0.66) | 5.46 (0.54) | 4.97 (0.87) | **6.57 (0.17)** | 5.00 (1.18) | 4.46 (0.47) | <0.001 |

Table IV. Mixed Effects Models to Assess the Trend in the Teacher Domain Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Teacher Domain Model 0** | | | **Teacher Domain Model 1** | | | **Teacher Domain Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.166 | -0.055 – 0.388 | 0.140 | -3.289 | -4.972 – -1.606 | **<0.001** | -3.293 | -5.160 – -1.426 | **0.001** |
| Year - 2021 | -0.162 | -0.276 – -0.047 | **0.006** | -0.147 | -0.260 – -0.035 | **0.011** | -0.147 | -0.260 – -0.034 | **0.011** |
| Average final grade | | | | 0.209 | 0.108 – 0.310 | **<0.001** | 0.219 | 0.105 – 0.333 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.015 | -0.163 – 0.192 | 0.871 |
| Academic year [2] | | | | | | | -0.314 | -0.979 – 0.351 | 0.349 |
| Academic year [3] | | | | | | | -0.447 | -1.092 – 0.199 | 0.172 |
| Academic year [4] | | | | | | | 0.080 | -0.535 – 0.694 | 0.797 |
| Academic year [5] | | | | | | | -0.046 | -0.653 – 0.561 | 0.880 |
| Academic year [6] | | | | | | | -0.935 | -1.764 – -0.105 | **0.028** |
| Random Effects | | | | | | | | | |
| $\sigma^2$ | 0.49 | | | 0.47 | | | 0.47 | | |
| $\tau_{00}$ | 0.51 $_{uc\_sem}$ | | | 0.42 $_{uc\_sem}$ | | | 0.38 $_{uc\_sem}$ | | |
| ICC | 0.51 | | | 0.47 | | | 0.45 | | |
| N | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | |
| Observations | 218 | | | 218 | | | 218 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.017 / 0.517 | | | 0.115 / 0.533 | | | 0.187 / 0.551 | | |

Table V. Mixed Effects Models to Assess the Trend in the UC Domain Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **UC Domain Model 0** | | | **UC Domain Model 1** | | | **UC Domain Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.004 | -0.223 – 0.231 | 0.975 | -3.608 | -5.108 – -2.108 | **<0.001** | -4.922 | -6.466 – -3.379 | **<0.001** |
| Year - 2021 | -0.024 | -0.105 – 0.057 | 0.555 | -0.009 | -0.088 – 0.069 | 0.813 | -0.007 | -0.086 – 0.072 | 0.869 |
| Average final grade | | | | 0.218 | 0.129 – 0.308 | **<0.001** | 0.260 | 0.167 – 0.353 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.437 | 0.266 – 0.608 | **<0.001** |
| Academic year [2] | | | | | | | 0.182 | -0.466 – 0.830 | 0.577 |
| Academic year [3] | | | | | | | 0.038 | -0.592 – 0.668 | 0.905 |
| Academic year [4] | | | | | | | 0.389 | -0.201 – 0.980 | 0.192 |
| Academic year [5] | | | | | | | 0.284 | -0.301 – 0.870 | 0.335 |
| Academic year [6] | | | | | | | -0.153 | -0.944 – 0.639 | 0.701 |
| Random Effects | | | | | | | | | |
| $\sigma^2$ | 0.25 | | | 0.23 | | | 0.23 | | |
| $\tau_{00}$ | 0.76 $_{uc\_sem}$ | | | 0.64 $_{uc\_sem}$ | | | 0.44 $_{uc\_sem}$ | | |
| ICC | 0.76 | | | 0.74 | | | 0.66 | | |
| N | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | |
| Observations | 218 | | | 218 | | | 218 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.000 / 0.755 | | | 0.111 / 0.765 | | | 0.319 / 0.765 | | |

Table VI. Mixed-Effects Models to Assess the Trend in the Difficulty Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Difficulty Model 0** | | | **Difficulty Model 1** | | | **Difficulty Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.263 | 5.104 – 5.423 | **<0.001** | 8.754 | 7.943 – 9.564 | **<0.001** | 8.218 | 7.384 – 9.052 | **<0.001** |
| Year - 2021 | -0.003 | -0.046 – 0.040 | 0.900 | -0.017 | -0.056 – 0.022 | 0.394 | -0.017 | -0.057 – 0.022 | 0.392 |
| Average final grade | | | | -0.211 | -0.259 – -0.163 | **<0.001** | -0.212 | -0.261 – -0.163 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.256 | 0.153 – 0.359 | **<0.001** |
| Academic year [2] | | | | | | | 0.261 | -0.134 – 0.655 | 0.191 |
| Academic year [3] | | | | | | | 0.365 | -0.019 – 0.749 | 0.062 |
| Academic year [4] | | | | | | | 0.230 | -0.127 – 0.586 | 0.203 |
| Academic year [5] | | | | | | | 0.452 | 0.099 – 0.806 | **0.013** |
| Academic year [6] | | | | | | | 0.401 | -0.075 – 0.877 | 0.097 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.07 | | | 0.06 | | | 0.06 | | |
| $\tau_{00}$ | 0.42 $_{uc\_sem}$ | | | 0.25 $_{uc\_sem}$ | | | 0.17 $_{uc\_sem}$ | | |
| ICC | 0.86 | | | 0.82 | | | 0.75 | | |
| N | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | |
| Observations | 219 | | | 219 | | | 219 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.000 / 0.859 | | | 0.245 / 0.860 | | | 0.489 / 0.872 | | |

## Figures

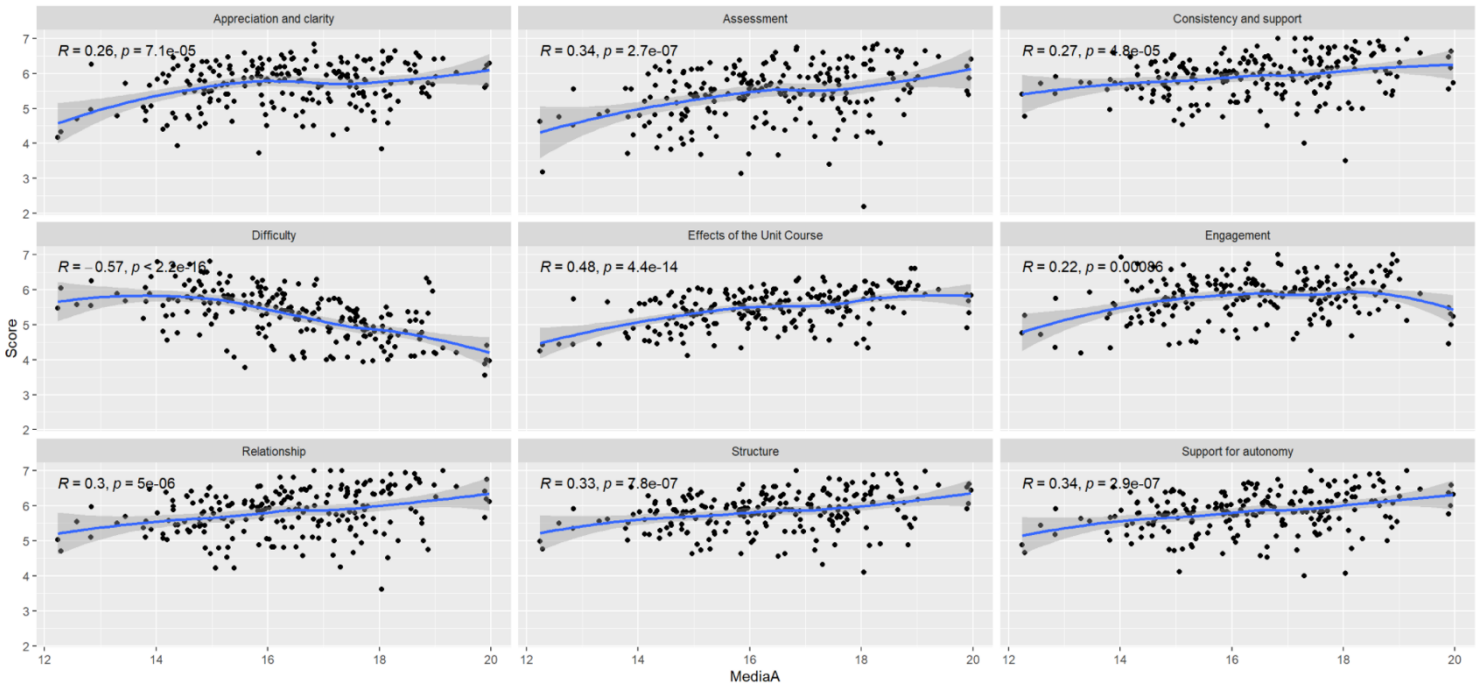



Figure 1. Relationship Between Evaluation Domains and Average Final Grade

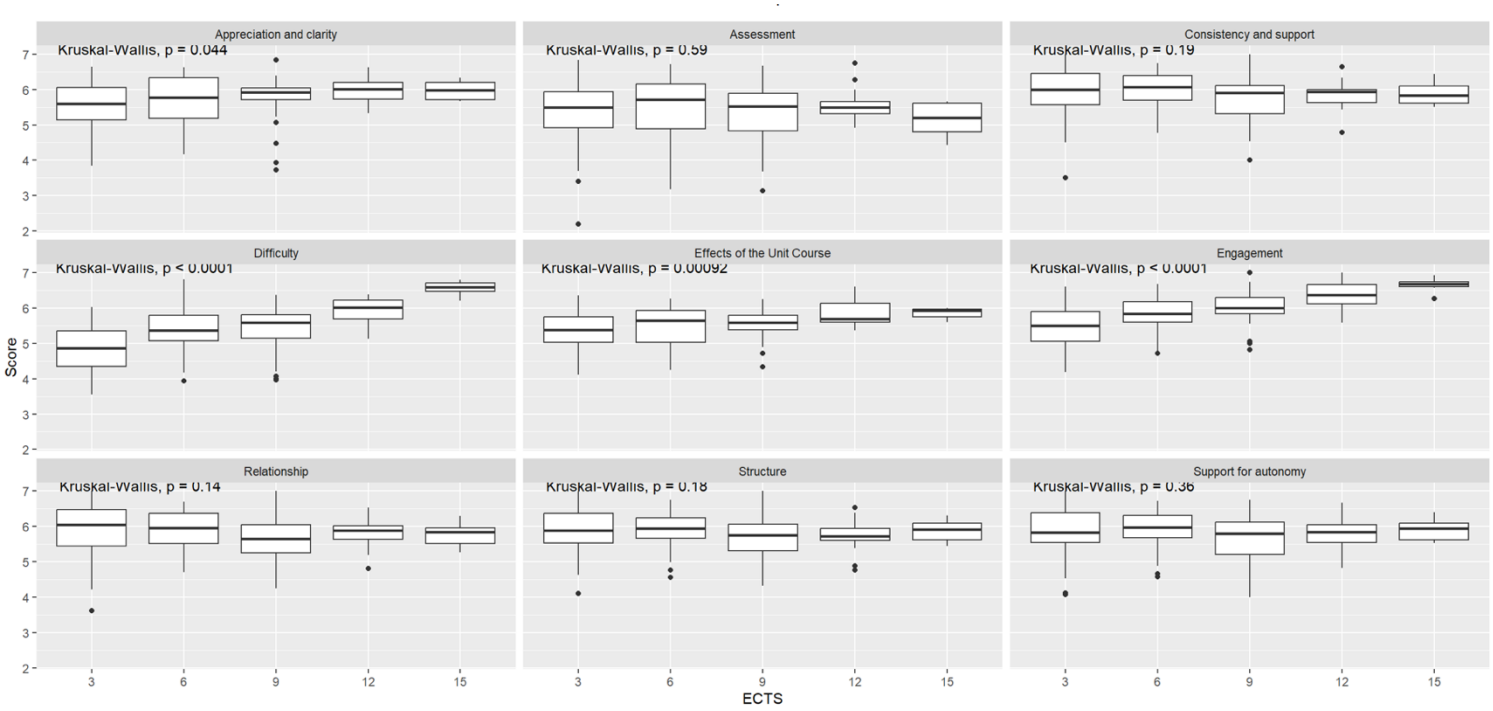



Figure 2. Relationship Between Evaluation Domains and Number of ECTS

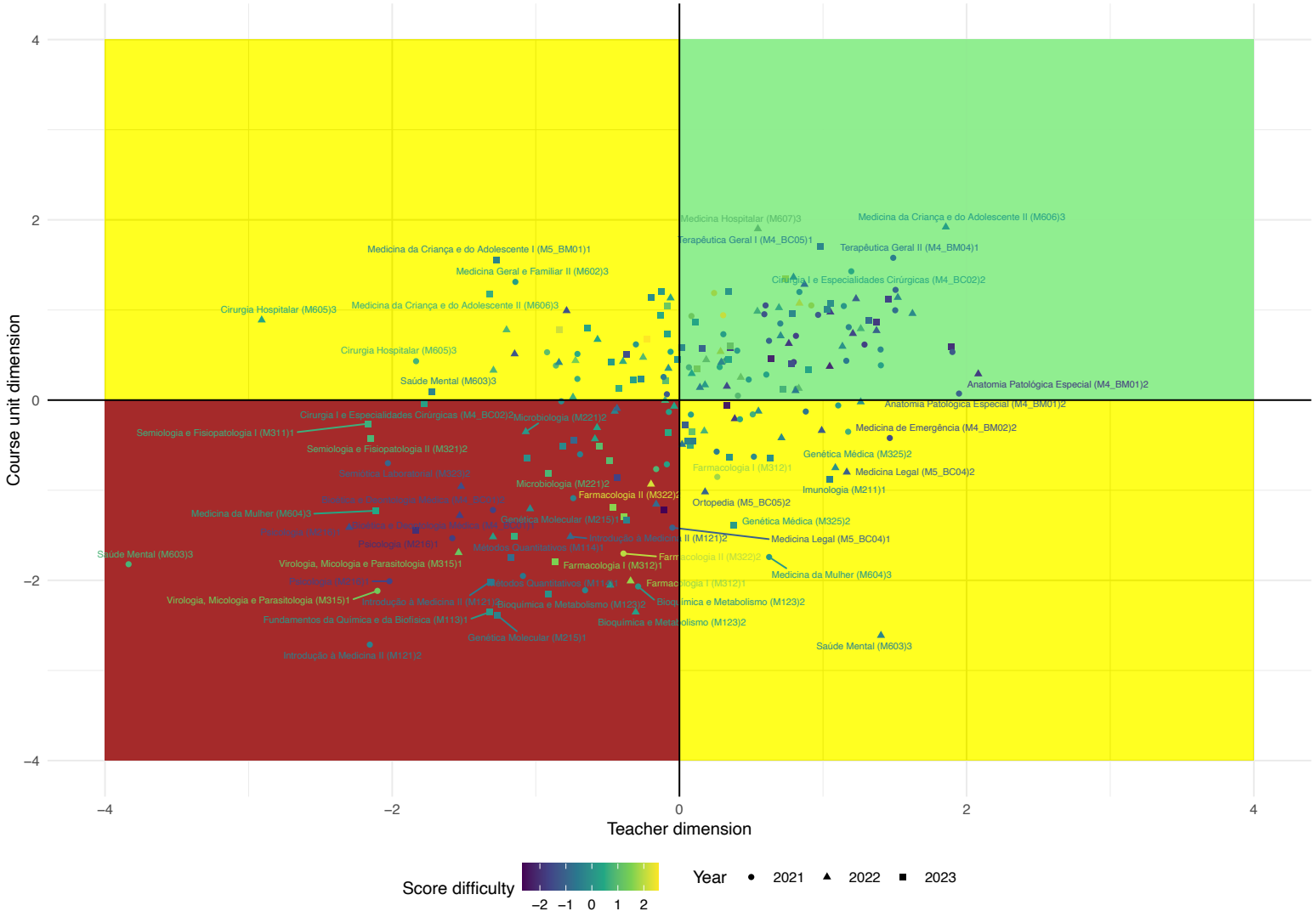



Figure 3. Scatter plot of UC across the Summary Domains of Teacher and Unit Course
Colour: Indicating the Summary Domain of Difficulty
Symbol: Representing the academic year

## Appendix

According to raw associations (Table ApI), the evaluation of Support for Autonomy declined by approximately 0.074 points per year ($p = 0.026$), a trend that remained statistically significant in adjusted models ($p = 0.045$). The association with average final grade was positive ($\beta = 0.128$; $p < 0.001$), while the number of ECTS showed no relevant effect ($p = 0.744$). The ICC of 0.52 indicates considerable consistency across UCs over time.

For Consistency and Help (Table ApII), a similar downward trend was observed (−0.097/year; $p = 0.011$), which persisted after adjusting for average final grade ($\beta = -0.089$; $p = 0.017$). Approval remained positively associated ($\beta = 0.108$; $p < 0.001$), but ECTS again showed no significant influence ($p = 0.650$). The ICC of 0.35 reflects moderate inter-UC variability.

Structure (Table ApIII) also saw a significant decline (−0.081/year; $p = 0.009$), slightly mitigated after adjustment ($\beta = -0.074$; $p = 0.016$). Average final grade correlated positively ($\beta = 0.108$; $p < 0.001$), with no notable ECTS effect ($p = 0.643$). With an ICC of 0.49, perceptions of this subdomain remained relatively stable between UCs.

In the case of Relationship (Table ApIV), evaluations dropped by 0.091 annually ($p = 0.020$). Adjusted estimates confirm this trend ($\beta = -0.082$; $p = 0.032$), alongside a positive link to average final grade ($\beta = 0.134$; $p < 0.001$). The number of ECTS was not a contributing factor ($p = 0.970$). The ICC of 0.44 suggests moderate consistency across courses.

Regarding Engagement (Table ApV), no significant variation over time was identified ($\beta = -0.003$; $p = 0.908$). Neither average final grade ($\beta = 0.003$; $p = 0.910$) nor ECTS ($\beta = 0.005$; $p = 0.829$) had meaningful associations. Nonetheless, an ICC of 0.63 points to strong inter-UC consistency.

As for Appreciation and Clarity (Table ApVI), there was a small but significant yearly reduction (−0.068; $p = 0.006$), attenuated after adjustments ($\beta = -0.062$; $p = 0.011$). Average final grade had a clear positive effect ($\beta = 0.096$; $p < 0.001$) and ECTS also emerged as significant ($\beta = 0.153$; $p = 0.010$). The ICC reached 0.75, indicating high stability across UCs.

The Assessment subdomain (Table ApVII) followed a similar pattern, with a notable decrease of 0.089/year ($p = 0.014$). After accounting for approval, the decline softened ($\beta =$

−0.077; p = 0.028), while average final grade remained a strong predictor (β = 0.176; p < 0.001). The number of ECTS coefficient lacked significance (p = 0.580) and the ICC was 0.68.

Lastly, UC Effects (Table ApVIII) showed no significant change over time (β = 0.013; p = 0.544), nor after adjusting for average final grade or ECTS. However, the ICC of 0.59 still reflects notable consistency in student perceptions across UCs.

## Appendix Tables

Table ApI. Mixed Effects Models to Assess the Trend for Support for Autonomy Over Academic Years
Model 1: Adjusted for average final grade
Model 2: Adjusted for average final grade, ECTS and academic year

| | **Support for Autonomy Model 0** | | | **Support for Autonomy Model 1** | | | **Support for Autonomy Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.915 | 5.789 – 6.041 | **<0.001** | 3.801 | 2.854 – 4.748 | **<0.001** | 3.737 | 2.712 – 4.763 | **<0.001** |
| Year – 2021 | -0.074 | -0.138 – -0.009 | **0.026** | -0.065 | -0.129 – -0.002 | **0.045** | -0.065 | -0.128 – -0.001 | **0.045** |
| Average final grade | | | | 0.128 | 0.071 – 0.184 | **<0.001** | 0.129 | 0.071 – 0.188 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.016 | -0.081 – 0.113 | 0.744 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.16 | | | 0.15 | | | 0.15 | | |
| $\tau_{00}$ | 0.17 $_{uc\_sem}$ | | | 0.14 $_{uc\_sem}$ | | | 0.14 $_{uc\_sem}$ | | |
| ICC | 0.52 | | | 0.48 | | | 0.48 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 221 | | | 221 | | | 221 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.011 / 0.522 | | | 0.125 / 0.542 | | | 0.123 / 0.545 | | |

Table ApII. Mixed Effects Models to Assess the Trend for Consistency and Help Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Consistency and Help Model 0** | | | **Consistency and Help Model 1** | | | **Consistency and Help Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 6.019 | 5.893 – 6.145 | **<0.001** | 4.223 | 3.257 – 5.190 | **<0.001** | 4.313 | 3.267 – 5.360 | **<0.001** |
| Year – 2021 | -0.097 | -0.171 – -0.023 | **0.011** | -0.089 | -0.163 – -0.016 | **0.017** | -0.090 | -0.163 – -0.016 | **0.017** |
| Average final grade | | | | 0.108 | 0.050 – 0.166 | **<0.001** | 0.106 | 0.046 – 0.165 | **0.001** |
| (ECTS - 3)/3 | | | | | | | -0.021 | -0.115 – 0.072 | 0.650 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.21 | | | 0.20 | | | 0.20 | | |
| $\tau_{00}$ | 0.13 $_{uc\_sem}$ | | | 0.11 $_{uc\_sem}$ | | | 0.11 $_{uc\_sem}$ | | |
| ICC | 0.38 | | | 0.35 | | | 0.35 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 221 | | | 221 | | | 221 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.018 / 0.390 | | | 0.097 / 0.412 | | | 0.099 / 0.417 | | |

Table ApIII. Mixed Effects Models to Assess the Trend for Structure Over Academic Years
Model 1: Adjusted for average final grade
Model 2: Adjusted for average final grade, ECTS and academic year

| | **Structure Model 0** | | | **Structure Model 1** | | | **Structure Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.926 | 5.807 – 6.045 | **<0.001** | 4.139 | 3.243 – 5.036 | **<0.001** | 4.227 | 3.256 – 5.197 | **<0.001** |
| Year – 2021 | -0.081 | -0.141 – -0.021 | **0.009** | -0.074 | -0.133 – -0.014 | **0.016** | -0.074 | -0.133 – -0.014 | **0.015** |
| Average final grade | | | | 0.108 | 0.054 – 0.162 | **<0.001** | 0.105 | 0.050 – 0.160 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | -0.021 | -0.113 – 0.070 | 0.643 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.14 | | | 0.13 | | | 0.13 | | |
| $\tau_{00}$ | 0.15 $_{uc\_sem}$ | | | 0.13 $_{uc\_sem}$ | | | 0.13 $_{uc\_sem}$ | | |
| ICC | 0.53 | | | 0.49 | | | 0.49 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 221 | | | 221 | | | 221 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.015 / 0.539 | | | 0.107 / 0.543 | | | 0.110 / 0.547 | | |

Table ApIV. Mixed Effects Models to Assess the Trend for Relationship Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Relationship Model 0** | | | **Relationship Model 1** | | | **Relationship Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.922 | 5.782 – 6.062 | **<0.001** | 3.700 | 2.626 – 4.773 | **<0.001** | 3.709 | 2.525 – 4.893 | **<0.001** |
| Year – 2021 | -0.091 | -0.167 – -0.015 | **0.020** | -0.082 | -0.157 – -0.007 | **0.032** | -0.082 | -0.157 – -0.007 | **0.032** |
| Average final grade | | | | 0.134 | 0.070 – 0.199 | **<0.001** | 0.134 | 0.067 – 0.201 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | -0.002 | -0.112 – 0.108 | 0.970 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.22 | | | 0.21 | | | 0.21 | | |
| $\tau_{00}$ | 0.19 $_{uc\_sem}$ | | | 0.16 $_{uc\_sem}$ | | | 0.16 $_{uc\_sem}$ | | |
| ICC | 0.46 | | | 0.43 | | | 0.44 | | |
| N | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | | 73 $_{uc\_sem}$ | | |
| Observations | 218 | | | 218 | | | 218 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.014 / 0.470 | | | 0.113 / 0.495 | | | 0.112 / 0.499 | | |

Table ApV. Mixed Effects Models to Assess the Trend for Engagement Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Engagement Model 0** | | | **Engagement Model 1** | | | **Engagement Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.792 | 5.663 – 5.922 | **<0.001** | 4.529 | 3.668 – 5.389 | **<0.001** | 3.282 | 2.476 – 4.089 | **<0.001** |
| Year – 2021 | -0.003 | -0.047 – 0.041 | 0.908 | 0.003 | -0.041 – 0.046 | 0.910 | 0.005 | -0.039 – 0.048 | 0.829 |
| Average final grade | | | | 0.076 | 0.025 – 0.128 | **0.004** | 0.110 | 0.065 – 0.156 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.333 | 0.249 – 0.416 | **<0.001** |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.07 | | | 0.07 | | | 0.07 | | |
| $\tau_{00}$ | 0.26 $_{uc\_sem}$ | | | 0.25 $_{uc\_sem}$ | | | 0.12 $_{uc\_sem}$ | | |
| ICC | 0.78 | | | 0.78 | | | 0.63 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 222 | | | 222 | | | 222 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.000 / 0.779 | | | 0.041 / 0.785 | | | 0.401 / 0.776 | | |

Table ApVI. Mixed Effects Models to Assess the Trend for Appreciation and Clarity Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Appreciation and Clarity Model 0** | | | **Appreciation and Clarity Model 1** | | | **Appreciation and Clarity Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.755 | 5.615 – 5.895 | **<0.001** | 4.159 | 3.236 – 5.082 | **<0.001** | 3.648 | 2.665 – 4.631 | **<0.001** |
| Year – 2021 | -0.068 | -0.116 – -0.020 | **0.006** | -0.062 | -0.109 – -0.015 | **0.011** | -0.061 | -0.108 – -0.014 | **0.012** |
| Average final grade | | | | 0.096 | 0.041 – 0.151 | **0.001** | 0.108 | 0.053 – 0.163 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.153 | 0.038 – 0.269 | **0.010** |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.09 | | | 0.08 | | | 0.08 | | |
| $\tau_{00}$ | 0.30 $_{uc\_sem}$ | | | 0.28 $_{uc\_sem}$ | | | 0.25 $_{uc\_sem}$ | | |
| ICC | 0.78 | | | 0.77 | | | 0.75 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 222 | | | 222 | | | 222 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.008 / 0.778 | | | 0.064 / 0.783 | | | 0.119 / 0.780 | | |

Table ApVII. Mixed Effects Models to Assess the Trend for Assessment Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **Assessment Model 0** | | | **Assessment Model 1** | | | **Assessment Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.524 | 5.346 – 5.702 | **<0.001** | 2.606 | 1.367 – 3.844 | **<0.001** | 2.465 | 1.124 – 3.805 | **<0.001** |
| Year – 2021 | -0.089 | -0.160 – -0.018 | **0.014** | -0.077 | -0.146 – -0.009 | **0.028** | -0.077 | -0.146 – -0.008 | **0.028** |
| Average final grade | | | | 0.176 | 0.102 – 0.250 | **<0.001** | 0.180 | 0.104 – 0.255 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.041 | -0.105 – 0.187 | 0.580 |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.19 | | | 0.18 | | | 0.18 | | |
| $\tau_{00}$ | 0.44 $_{uc\_sem}$ | | | 0.38 $_{uc\_sem}$ | | | 0.38 $_{uc\_sem}$ | | |
| ICC | 0.70 | | | 0.68 | | | 0.68 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 222 | | | 222 | | | 222 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.008 / 0.703 | | | 0.122 / 0.719 | | | 0.120 / 0.721 | | |

Table ApVIII. Mixed Effects Models to Assess the Trend for UC Effects Over Academic Years

Model 1: Adjusted for average final grade

Model 2: Adjusted for average final grade, ECTS and academic year

| | **UC Effects Model 0** | | | **UC Effects Model 1** | | | **UC Effects Model 2** | | |
|---|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* | *Estimates* | *CI* | *p* |
| (Intercept) | 5.472 | 5.357 – 5.586 | **<0.001** | 3.057 | 2.311 – 3.803 | **<0.001** | 2.293 | 1.547 – 3.039 | **<0.001** |
| Year – 2021 | 0.013 | -0.030 – 0.057 | 0.544 | 0.023 | -0.019 – 0.065 | 0.275 | 0.025 | -0.017 – 0.066 | 0.246 |
| Average final grade | | | | 0.146 | 0.101 – 0.190 | **<0.001** | 0.169 | 0.127 – 0.211 | **<0.001** |
| (ECTS - 3)/3 | | | | | | | 0.187 | 0.112 – 0.263 | **<0.001** |
| **Random Effects** | | | | | | | | | |
| $\sigma^2$ | 0.07 | | | 0.07 | | | 0.07 | | |
| $\tau_{00}$ | 0.19 $_{uc\_sem}$ | | | 0.14 $_{uc\_sem}$ | | | 0.09 $_{uc\_sem}$ | | |
| ICC | 0.73 | | | 0.68 | | | 0.59 | | |
| N | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | | 74 $_{uc\_sem}$ | | |
| Observations | 222 | | | 222 | | | 222 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.000 / 0.727 | | | 0.197 / 0.740 | | | 0.341 / 0.730 | | |

## References

1. Berk RA. Using the 360 degrees multisource feedback model to evaluate teaching and professionalism. Med Teach. 2009;31(12):1073-80.

2. Schiekirka S, Feufel MA, Herrmann-Lingen C, Raupach T. Evaluation in medical education: A topical review of target parameters, data collection tools and confounding factors. GMS German Medical Science. 2015;13:Doc15.

3. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Medical education. 2005;39(3):309-17.

4. Dunham L, Dekhtyar M, Gruener G, CichoskiKelly E, Deitz J, Elliott D, et al. Medical student perceptions of the learning environment in medical school change as students transition to clinical training in undergraduate medical school. Teaching and learning in medicine. 2017;29(4):383-91.

5. Khan M, Mourshed G, Tasneem J, Azad S, Afrin N. Phase wise evaluation of teachers by the students in Mymensingh Medical College, Bangladesh. Mymensingh Medical Journal: MMJ. 2023;32(4):906-10.

6. de Lemos MS, Menezes I, Teixeira PM, Queirós C. Construção e validação do Inquérito Pedagógico da Universidade do Porto. 2014.

7. Nora LM, Pouwels MV, Irons M. Expanding educators' contributions to continuous quality improvement of American Board of Medical Specialties maintenance of certification. Academic Medicine. 2016;91(1):16-9.

8. Willett TG, Marshall KC, Broudo M, Clarke M. It's about TIME: a general-purpose taxonomy of subjects in medical education. Medical education. 2008;42(4):432-8.

9. Berkhout JJ, Helmich E, Teunissen PW. The complex relationship between student, context and learning outcomes. Medical Education. 2016;50(2):164-6.

10. Issenberg SB, Scalese RJ. Multimedia computer-based measures of basic science and clinical skills. Medical education. 2004;38(5).

11. Genn J. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education–a unifying perspective. Medical teacher. 2001;23(5):445-54.

12. Till H. Climate studies: can students' perceptions of the ideal educational environment be of use for institutional planning and resource utilization? Medical Teacher. 2005;27(4):332-7.

13. Marshall RE. Measuring the medical school learning environment. Academic Medicine. 1978;53(2):98-104.

14. Uttl B, White CA, Gonzalez DW. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Studies in Educational Evaluation. 2017;54:22-42.

15. Young PW, editor Student-produced video of role-plays on topics in cell biology and biochemistry: A novel undergraduate group work exercise. Frontiers in Education; 2020: Frontiers Media SA.

16. Lizzio A, Wilson K, Simons R. University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. Studies in Higher education. 2002;27(1):27-52.

17. Kuhn K-A, Rundle-Thiele S. Curriculum alignment: Student perception of learning achievement measures. International Journal of Teaching and Learning in Higher Education. 2009;21(3):351-61.

18. Sechelski A, Bhattacharjee R, Reynolds A, Manis M, Elshabrawy HA, Zhao Y. Analysis of Student Perceptions of a Newly Developed Integrative System Course Model. Medical Science Educator. 2023;33(5):1061-71.

19. Khader AA, Obeidat FN, Shahin NA, Khouri NA, Kaddumi EG, Al Qa'qa S, et al. Medical Students' Perceptions of Pathology and a Proposed Curricular Integration with Histology: A Future Vision of Curricular Change. International Journal of Morphology. 2020;38(1).

20. Diogo PG, Moreira A, Coimbra A, Silva AC, Martins AN, Mendonça C, et al. Estudo sobre as condições pedagógicas das escolas médicas portuguesas: uma análise nacional sobre a satisfação estudantil, rácios estudante-tutor e número de admissões. Acta Medica Portuguesa. 2016;29(5):301-9.

21. Thomas M, Kettaneh A, Fain O. Evaluation of medical semiology teaching and teachers. Study of student satisfaction in the second cycle first year. Presse Medicale (Paris, France: 1983). 2002;31(25):1165-71.

22. Sobral DT. Cross-year peer tutoring experience in a medical school: conditions and outcomes for student tutors. Medical education. 2002;36(11):1064-70.

23. Fasinu PS, Wilborn TW. Pharmacology education in the medical curriculum: Challenges and opportunities for improvement. Pharmacology Research & Perspectives. 2024;12(1):e1178.

24. Pandit R, Poleij MC, Gerrits MA. Student Perception of Knowledge and Skills in Pharmacology and Pharmacotherapy in a Bachelor's Medical Curriculum. International Medical Education. 2023;2(3):206-18.

25. Harden RM, Laidlaw JM. Essential skills for a medical teacher: an introduction to teaching and learning in medicine: Elsevier Health Sciences; 2020.

26. Garrison DR, Kanuka H. Blended learning: Uncovering its transformative potential in higher education. The internet and higher education. 2004;7(2):95-105.

27. Hattie J, Timperley H. The power of feedback. Review of educational research. 2007;77(1):81-112.

28. Thomas A, Lubarsky S, Durning SJ, Young ME. Knowledge syntheses in medical education: demystifying scoping reviews. Academic Medicine. 2017;92(2):161-6.

29. Al-Azri H, Ratnapalan S. Problem-based learning in continuing medical education: review of randomized controlled trials. Canadian family physician. 2014;60(2):157-65.

30. Bucklin BA, Valley M, Welch C, Tran ZV, Lowenstein SR. Predictors of early faculty attrition at one Academic Medical Center. BMC Medical Education. 2014;14:1-7.

INSTITUTO DE **CIÊNCIAS BIOMÉDICAS ABEL SALAZAR**