BMC Medical Education

RESEARCH

Open Access

Trajectories of medical students' empathy nowadays: a longitudinal study using a comprehensive framework of empathy

Valerie Carrard[1], Céline Bourquin[1], Sylvie Berney[2], Setareh Ranjbar[3], Katja Schlegel[4], Jacques Gaume[5], Pierre-Alexandre Bart[6], Marianne Schmid Mast[7], Martin Preisig[3] and Alexandre Berney[1*]

Abstract

Background For more than a decade, the literature has been dominated by the notion that medical students may paradoxically lose their empathy during medical school. However, medical curricula have significantly evolved, and the question is whether this is still the case. The present study aimed to describe the trajectories of different dimensions of empathy from the beginning to the end of a six-year medical curriculum and explore the influence of different psychosocial and health-related factors.

Methods In an open cohort design, all medical students at the University of Lausanne (Switzerland) were invited to complete four waves of yearly online questionnaires. Cognitive, affective, and behavioral empathy were measured with three validated instruments, and emotion recognition was assessed with a performance test. For each measure, linear mixed models including an array of psychosocial and health-related potential covariates were modelled. Different temporal variance–covariance structures and nonlinear trajectories were tested.

Results The final sample included 3224 questionnaires completed by 1667 medical students. The cognitive and affective dimensions of empathy significantly increased in the first half of medical school, followed by a plateau, whereas behavioral empathy remained stable. For emotion recognition, a significant linear increase was observed. The only covariate with substantial influence was gender: students identifying as male presented similar trajectories of empathy and emotion recognition but with overall lower scores than students identifying as female or nonbinary.

Conclusions This study revealed significant increases in cognitive empathy, affective empathy, and emotion recognition. Developments in today's medical curricula may have contributed to the observed increase in empathy. Future multisite studies are warranted to identify the features of the educational environment that impact the trajectories of empathy during medical school.

Keywords Medical students, Empathy, Emotion recognition, Longitudinal study

*Correspondence:
Alexandre Berney
alexandre.berney@chuv.ch
Full list of author information is available at the end of the article

Introduction

Since the seminal work of Hojat and colleagues [1] in 2009, the decrease in empathy during medical school has been widely broadcasted and the belief that medical students "lose their empathy" during medical school has become well rooted in both the medical and medical education communities. However, after more than a decade of medical curriculum improvement worldwide regarding the relational and communicational aspects of patient encounters [2, 3], the time might have come to reexamine the trajectories of empathy during medical school. In 2020, a systematic review reported that, among the 24 cross-sectional and 6 longitudinal studies identified, 14 reported a significant decrease in empathy during medical school, whereas the remaining 16 reported an increase, stability, or mixed results [4]. Some authors have attributed these mixed findings to geo-sociocultural factors. A review indicated that US studies mostly reported small but significant decreases in empathy, whereas Far Eastern studies mostly reported small but significant increases in empathy [5]. Another explanation could be the instruments used to measure empathy. A meta-analysis indeed revealed that changes in empathy were significant only when they were measured with the Jefferson Scale of Physician Empathy (JSPE [6]) [7]. One study even reported opposing trajectories when different instruments were used. Smith et al. [8] indeed found the decline in empathy previously described when the JSPE was used but reported a significant increase in empathy when measured with the Questionnaire of Cognitive and Affective Empathy (QCAE [9]). Also, using task-based measures in addition to self-report questionnaires, their results revealed that medical students' ability to recognize others' emotional states and their sensitivity to facial expressions increased during medical school [8].

It is important to consider the well-recognized multidimensionality of empathy. Most authors agree that it encompasses at least a cognitive and an affective component, which are, respectively, the ability to recognize and understand others' emotions by taking their perspective for cognitive empathy and a resonance with or microcontagion of others' feelings for affective empathy [10]. Other authors add to these two dimensions a behavioral component of empathy, which is the ability to act accurately on the basis of one's understanding of others' emotions [11]. Using the Interpersonal Reactivity Index (IRI [12]), a multidimensional measure of empathy, Quince et al. [13] reported that different dimensions of empathy can follow different trajectories during medical school. They indeed reported a statistically significant decline in the affective dimension of empathy, whereas the cognitive dimension of empathy remained stable [13]. They also observed that this decline in affective empathy was only significant for male students, thus highlighting the importance of accounting for both the multidimensionality of empathy and the impact of potentially influencing factors.

Several factors have been shown to influence empathy and its trajectories, such as gender [1, 4, 14], specialty choice [1, 4, 14], social support [14], and psychological well-being [14–19]. To obtain a better understanding of empathy trajectories during medical school, these factors need to be controlled for. Moreover, identifying influencing factors would provide a better understanding of individual differences in empathy trajectories and may offer avenues for interventions to prevent empathy decline during medical school.

The aim of the present study was to describe the trajectories of different dimensions of empathy over the full period of medical school and explore the influence of different psychosocial and health factors on these trajectories. This study sought to build on the same strengths as previous work [8, 13] by adopting a comprehensive framework of empathy with validated self-report questionnaires of the cognitive, affective, and behavioral components of empathy as well as a test of emotion recognition to include a measure that does not rely on self-reports. Moreover, it complements existing literature by (1) covering an entire six-year curriculum, (2) including a large sample, (3) formally testing possible nonlinear trajectories of empathy, and (4) investigating the influence of psychosocial and health-related factors. Such studies revisiting empathy trajectories during medical school are currently needed to maybe start deconstructing the notion of medical students' empathy loss. Indeed, medical curricula constantly evolve, and the efforts made during the last two decades to improve communication skills training [2, 3] might hopefully have impacted the development of empathy during medical school.

Method

Design

This longitudinal study used data collected in an open cohort design for the ETMED-L project [20]. In each of the four data collection waves (March 2021, November 2021, November 2022, and November 2023), medical students from all curriculum years (1 to 6) were included (except for the last data collection wave, in which first-year students were excluded to prioritize multiple participation). Students' longitudinal participation was tracked using their university email address. Then, data were organized according to curriculum years (instead of data collection waves) before applying Linear Mixed Models (LMMs). LMM is a statistical method that extends standard linear regression to account for both fixed effects, which represent overall population-level trends, and

random effects, which capture individual-level variability. In the context of repeated measures, LMMs are particularly useful as they model the correlations between repeated observations within the same subject, allowing for a more accurate analysis of within-subject and between-subject differences over time. Moreover, LMMs were chosen because they effectively accommodate limited and discontinuous data, providing a reliable way to estimate the longitudinal trajectories of empathy across the complete six years of medical education, even if assessments were conducted at a maximum of four time points for each participant [21].

Data collection

At each of the four data collection waves, all medical students matriculated at the University of Lausanne in Switzerland, excluding external exchange students, were invited to complete an online questionnaire investigating mental health and interpersonal competence [22]. Completing the online questionnaire took about one hour, and students were compensated with 50 CHF (~ 50 USD) for each completed questionnaire. The Cantonal Research Ethics Committee—Vaud (Commission cantonale d'éthique de la recherche sur l'être humain – Vaud [CER-VD]) approved the project (project number 2020–02474), and all participants provided written informed consent. At the end of the questionnaire, contact information for mental health help services was provided if students felt the need to seek support.

Measures

Empathy and emotion recognition

To adopt a more comprehensive framework of empathy, we used five measures from three validated instruments as well as an emotion recognition test described in more details in the published study protocol [20]. First, we assessed empathy with the widely used Jefferson Scale of Physician Empathy—Student Version (JSPE-S [6]), which was developed to measure medical students' orientations or attitudes toward empathic relationships in the context of patient care [6]. Second, the validated French version of the Questionnaire of Cognitive and Affective Empathy (QCAE [9, 23]) was used as a more multidimensional measure of empathy assessing separately cognitive and affective empathy. For assessing behavioural empathy, the observation of interactions with patients would have been ideal, but these kinds of data would have been difficult to obtain. After careful examination of existing self-reported instruments, we found that the validated French version of the Ability to Modify Self-Presentation Scale (AMSP [24, 25]) could be a proxy to measure a behavioral dimension of empathy, as it assesses the ability to modify one's behaviours according to the social situation at hand. Finally, we used the Geneva Emotion Recognition Test – Short Form (GERT-S [26]) to measure emotion recognition accuracy. Indeed, empathy has traditionally been assessed through self-report questionnaires, but there are also well-validated performance-based tasks that evaluate the ability to recognize emotions in individuals depicted in pictures or short videos [27]. The GERT-S presents 42 video clips of actors expressing one out of 14 emotions while saying pseudolinguistic sentences. The final score of the test is the number of emotions correctly recognized by the participants. These types of emotion recognition tasks have been found to have a significant, though modest, correlation with self-reported measures of both cognitive and affective empathy (Murphy and Lilienfeld, 2019), suggesting that while the ability to recognize others' emotions is linked to both understanding and resonating with others' emotions, they also capture aspects of emotional processing that go beyond these dimensions.

Psychosocial and health-related covariates

Four *sociological covariates* were examined: identifying as male (gender identity recoded as 1 for male and 2 for female or nonbinary), parental education (the number of parents with a college or university degree), relationship status (1 for having a partner and 0 for not), and social support. Social support was evaluated using the average score from two questions adapted from the Swiss Household Panel survey [28]: "If necessary, in your opinion, to what extent can someone provide you with practical help, this means concrete help or useful advice, if 0 means "not at all" and 10 "a great deal"?" for practical support and "To what extent can someone be available in case of need and show understanding, by talking with you for example, if 0 means "not at all" and 10 "a great deal"?" for emotional support.

In terms of *psychological covariates*, we assessed whether participants had consulted a psychotherapist in the past year (1 for yes, 0 for no) and their coping strategies. The French version of the coping section from the Euronet questionnaire [29, 30] was utilized to evaluate three types of coping strategies: emotion-focused, problem-focused, and help-seeking [31]. Moreover, mental health and burnout indicators were included. For mental health, the Center for Epidemiological Studies-Depression [32] was used to measure depression symptoms, and the State-Trait Anxiety Inventory [33] was used for anxiety symptoms. Additionally, the emotional exhaustion, cynicism, and academic efficacy (reversed) dimensions of burnout were assessed with the Maslach Burnout Inventory Student-Survey [34].

Lastly, two *health-related covariates* were assessed: physical activity (measured in hours per week) and self-reported health satisfaction ("Are you satisfied with your

health?" rated on a scale from 1 for very unsatisfied to 5 for very satisfied).

Statistical analysis

To investigate the longitudinal trajectories of empathy during medical school, separate LMMs were modelled for each empathy and emotion recognition measure, with curriculum year and covariates as fixed effects. Specifically, the fixed effects account for the systematic variation associated with changes across the curriculum years and control for potential confounding factors. Additionally, random intercepts were incorporated at the student level to account for the correlation between repeated measures within the same student while considering the inherent variability between students. Restricted maximum likelihood was utilized to obtain unbiased estimates of variance and covariance parameters [35, 36]. Furthermore, two different temporal variance–covariance structures (Autoregressive Covariance Structure of AR1 and Autoregressive/Moving Average Covariance Structure of ARMA1.1) were tested to potentially account for the temporal spillover of empathy and emotion recognition, with the best fitting model being selected via likelihood ratio tests. The nonlinear trajectories of empathy and emotion recognition were subsequently tested by including the time as quadratic or cubic. The best-fitting model was selected via likelihood ratio tests and presented as the final model. Standardized β values are reported for estimating effect sizes, with values ranging from 0.10 to 0.29 being considered as small, those from 0.30 to 0.49 as medium, and values of 0.50 or higher as large effect sizes [37].

To explore significant differences between specific time points, pairwise comparisons to each preceding year were additionally conducted with Holm-Bonferroni corrections to account for multiple comparisons. However, unlike LMMs, these comparisons do not account for the correlation between repeated measures within the same student and are thus a less accurate representation of longitudinal trajectories.

The highest missing rate at the item level was 0.59% (AMSP items). It has been shown that much gain from multiple imputation is unlikely when missing rates are lower than 5% [38]. As a result, if fewer than 20% of the items in an instrument were missing, those items were replaced with mean scores. However, if 20% or more of the items were missing, the total score for the instrument was deemed missing, and these missing values were then handled in the LMMs with full information maximum likelihood. All the analyses were run in Stata version 17 [39] and R version 4.2.2 [40], and p values < 0.05 were considered significant.

Results

Sample

Figure 1 presents the flow chart of the study participation. After exclusion of the students who gave a wrong answer to at least one of the two attention questions (e.g., "In order to check your attention, please answer 'Slightly agree' to this question."), scored implausibly low on the

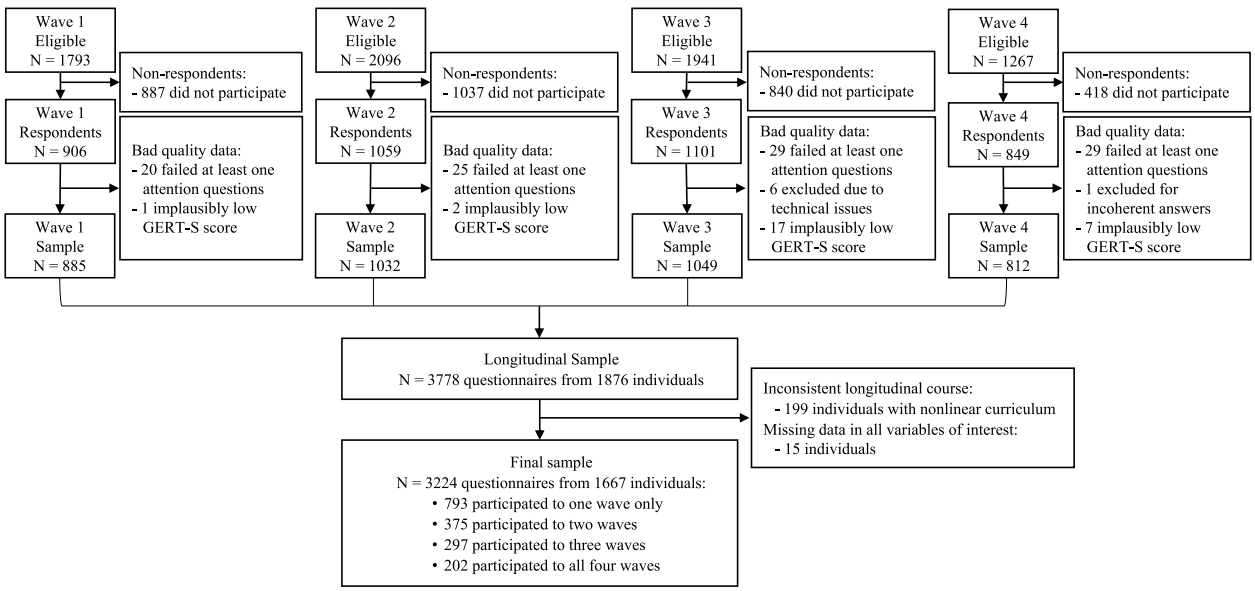

Fig. 1 Participation flow chart for each data collection wave. Note that in wave 4, first-year students were not eligible for participation. GERT-S = Geneva Emotion Recognition Test – Short form

GERT-S, or encountered technical issues, each data collection wave included respectively 885, 1032, 1049, and 812 answered questionnaires, which represented 49.36%, 57.56%, 58.51%, and 45.29% of each wave's eligible students. From those, we excluded 169 repeat students to have only linear curriculum trajectories as well as 15 students who had missing data on all the variables of interest across all the data collection waves. Consequently, the final sample included 1667 medical students who filled in a total of 3224 questionnaires.

The medical students in our final sample were between 17 and 49 years old, with a mean age of 21.80 years (standard deviation = 3.09) at first participation. Among them, 67.19% self-identified as female, 0.84% self-identified as nonbinary, and 31.97% as male, which aligns with the gender distribution typically seen in the Lausanne Medical School. Half of the medical students had a partner (52.37%), and the majority had one (21.66%) or two (53.57%) parents with higher education (college or university diploma). An important proportion of the students, 23.88%, indicated having consulted a psychotherapist in the last 12 months, which is more than twice as much as the corresponding rates in the general population of similar ages (9% for the 15–24 years old and 11% for the 25–34 years old [41]).

Trajectories of empathy and emotion recognition

The results of the LMMs modelling empathy trajectories from the first to the sixth year of medical school are displayed in Table 1. For the JSPE-S, the cognitive dimension of the QCAE, and the affective dimension of the QCAE, we observed a significant general increase over the curriculum year (*standardized* β [*std.* β] = 0.21, 0.05, and 0.06, respectively), taking a nonlinear concave shape (*std.* β = -0.16, -0.04, and -0.03, respectively). However, the pairwise comparisons to each preceding year displayed in Table 2 indicate that the shape of the empathy trajectories might be one of significant increase followed by a plateau. Indeed, we observed significant increases in JSPE-S scores from year 1 up to year 4, followed by a plateau, with no significant changes from year 4 to 6. For both the cognitive and affective dimensions of the QCAE, we observe a single significant increase from year 2 to year 3, followed also by a plateau with no significant changes between year 3 and 6. For the behavioral dimension of empathy measured with the AMSP, the curriculum years had no significant impact, indicating a stable trajectory. Finally, the emotion recognition test (GERT-S) results followed a linear trajectory of significant increase (*std.* β = 0.20), and pairwise comparisons to each previous year revealed significant increases from year 2 up to year 5. The trajectories' effect sizes are small for the JSPE-S and the GERT-S and even smaller for the cognitive and affective dimensions of the QCAE.

Influence of psychosocial and health-related covariates

As shown in Table 1, several covariates significantly influenced different measures of empathy and emotion recognition. Nevertheless, identifying as male was the only covariate with an absolute average effect size large enough to be considered small (absolute average β = 0.13). Compared with the students identifying as female or nonbinary, the students identifying as male had lower scores on the JSPE-S, QCAE (both cognitive and affective), and GERT-S, whereas the reverse was observed for the AMSP, with students identifying as male having higher scores than students identifying as female or nonbinary. Note that the interaction effect between gender identification and curriculum year was also investigated and was never significant, indicating that even if students' level of empathy varied according to their gender identification, the slopes of the trajectories were similar. All other covariates had absolute average effect sizes that were too low to even be considered as small.

Discussion

This study aimed to offer an up-to-date insight into empathy trajectories in today's undergraduate medical education. Its results do not confirm the widely broadcasted decline in empathy during medical school. Among the five included instruments measuring different dimensions of empathy and emotion recognition, none followed a decreasing trajectory. On the contrary, this study shows a general enhancement in cognitive and affective empathy, especially in the first curriculum years, and a steady improvement in emotion recognition abilities. Behavioral empathy, on the other hand, remained relatively stable during medical school.

The present study is not the first to show an improvement in empathy during medical school [4, 8], but its results contradict those reported in previous studies using the JSPE [7]. These differences might be due to geo-sociocultural factors such as differences in the conceptualization of empathy, different ways to promote empathy, and differences in the educational environment, resources and logistics within a given medical school. Nevertheless, in the present study, an increase was observed across different measures and different dimensions of empathy. Importantly, the measure that shows the most consistent increase across time is the emotion recognition test (GERT-S). This test does not rely on self-reports and could be considered a more reliable estimate of the students' actual abilities, which strengthens our findings that the students tended to become more empathic during

Table 1 Linear mixed models of empathy trajectories including psychosocial and health-related covariates

	JSPE-S			QCAE-Cognitive			QCAE-Affective			AMSP			GERT-S		
Predictors	***std. β***	***B***	***p***	***std. β***	***B***	***p***	***std. β***	***B***	***p***	***std. β***	***B***	***p***	***std. β***	***B***	***p***
(Intercept)	.17	87.11	**<.001**	.06	50.16	**<.001**	.02	23.61	**<.001**	.03	18.73	**<.001**	-.04	28.35	**<.001**
Curriculum year	.21	4.72	**<.001**	.05	0.95	**<.001**	.06	0.64	**.001**	-.03	-0.09	.110	.20	0.51	**<.001**
Curriculum year^2	-.16	-0.52	**<.001**	-.04	-0.11	**.002**	-.03	-0.06	**.019**						
Identifying as male	-.08	-1.50	**.001**	-.11	-1.55	**<.001**	-.23	-2.77	**<.001**	.06	0.68	**.008**	-.18	-1.61	**<.001**
Parental Education	-.01	-0.10	.658	.01	0.11	.555	.02	0.14	.309	.03	0.15	.250	.07	0.38	**.001**
Having a partner	.04	0.77	**.011**	.01	0.09	.684	.02	0.18	.259	.05	0.49	**.004**	.03	0.23	.132
Social support	.08	0.38	**<.001**	.04	0.15	**.012**	.06	0.18	**<.001**	.02	0.05	.265	.08	0.17	**<.001**
Psychotherapist consultation	.04	0.80	**.024**	.05	0.76	**.002**	.02	0.26	.170	.03	0.41	**.042**	-.01	-0.12	.501
Emotion-focused coping	.03	0.07	.148	-.02	-0.03	.434	.17	0.24	**<.001**	.00	0.00	.990	.02	0.02	.351
Problem-focused coping	.05	0.27	**.001**	.13	0.49	**<.001**	.02	0.07	.111	.07	0.21	**<.001**	.01	0.03	.443
Help-seeking coping	.04	0.13	**.027**	.05	0.12	**.005**	.11	0.21	**<.001**	-.01	-0.01	.763	-.04	-0.06	.050
Depressive symptoms	.03	0.02	.287	.06	0.03	**.014**	.08	0.04	**<.001**	.00	0.00	.931	.00	0.00	.943
Anxiety symptoms	-.05	-0.04	.069	-.07	-0.04	**.008**	.11	0.05	**<.001**	-.16	-0.07	**<.001**	-.02	-0.01	.445
Burnout-Emotional exhaustion	.06	0.10	**.009**	.01	0.02	.516	.01	0.01	.482	.06	0.06	**.012**	-.02	-0.01	.449
Burnout-Cynicism	.00	0.01	.829	.00	-0.01	.817	-.03	-0.03	.156	.04	0.05	**.037**	.00	0.00	.924
Burnout-Academic efficacy	.08	0.16	**<.001**	.09	0.13	**<.001**	.02	0.03	.214	.15	0.16	**<.001**	-.04	-0.03	.113
Physical activity	.01	0.04	.464	-.02	-0.05	.283	-.02	-0.03	.311	.06	0.12	**<.001**	.00	0.00	.880
Satisfaction with health	.00	-0.04	.769	-.03	-0.19	.070	.04	0.25	**.002**	-.03	-0.15	.071	-.01	-0.05	.519
$N_{individuals}$	1658			1664			1664			1658			1662		
$N_{observations}$	3205			3213			3213			3204			3207		
σ^2	33.28			14.87			8.56			10.79			8.05		
$\tau_{00\ individuals}$	36.69			26.34			14.26			12.66			8.79		
ICC	.52			.64			.62			.54			.52		
Marginal R^2	.12			.07			.23			.08			.09		
Conditional R^2	.58			.66			.71			.58			.56		
AICc	22,066.31			20,015.28			18,203.28			18,422.02			17,664.37		

Significant p-values are indicated in bold

Std Standardized, *JSPE-S* Jefferson Scale of Physician Empathy Student version, *QCAE* Questionnaire of Cognitive and Affective Empathy, *AMSP* Ability to Monitor Self-Presentation, *GERT-S* Geneva Emotion Recognition Test Short. R^2 indicates the percentage of variation accounted for by the fixed component of the model, conditional R^2 represent the percentage of variation explained by both the fixed and random components, intraclass correlation coefficient (ICC) reflects the proportion of total variance attributed to differences between students. Standardized βs between 0.10–0.29 are considered small, those between 0.30–0.49 are considered medium, and those 0.50 or greater are considered large effect sizes [37]

Table 2 Descriptive statistics of students' empathy per curriculum year and comparisons to previous years

	N	%Missing	Mean	SD	Min	Max	Comparison to previous year: *z* (SE) *p*-value		
JSPE-S									
Year 1	574	0.69	101.71	9.98	65	126			
Year 2	448	0.22	105.26	8.40	74	125	7.67	(0.43)	**< .001**
Year 3	579	0.69	107.76	8.28	79	127	5.85	(0.39)	**< .001**
Year 4	545	0.55	108.85	7.92	67	130	2.44	(0.37)	**.044**
Year 5	602	0.17	108.48	7.92	77	126	-0.70	(0.36)	.499
Year 6	458	1.08	108.55	8.58	76	127	-1.15	(0.39)	.499
QCAE-Cognitive									
Year 1	578	0.00	57.36	6.69	35	76			
Year 2	449	0.00	57.56	7.00	36	76	1.30	(0.30)	.585
Year 3	583	0.00	58.79	6.57	34	76	3.14	(0.26)	**.008**
Year 4	548	0.00	58.38	7.19	33	76	-0.79	(0.25)	.585
Year 5	603	0.00	58.85	6.63	35	76	1.21	(0.24)	.585
Year 6	463	0.00	58.74	6.52	36	76	-1.48	(0.26)	.553
QCAE-Affective									
Year 1	578	0.00	33.86	5.82	15	47			
Year 2	449	0.00	33.90	5.54	17	48	1.37	(0.23)	.679
Year 3	583	0.00	34.58	5.57	16	48	2.66	(0.20)	**.039**
Year 4	548	0.00	34.67	5.49	20	48	0.35	(0.19)	1.000
Year 5	603	0.00	34.92	5.44	15	48	0.10	(0.19)	1.000
Year 6	463	0.00	34.87	5.34	18	48	0.17	(0.20)	1.000
AMSP									
Year 1	574	0.69	23.51	5.27	10	35			
Year 2	448	0.22	22.80	5.29	5	35	-3.68	(0.24)	**.001**
Year 3	579	0.69	23.03	5.06	5	35	0.90	(0.22)	.989
Year 4	544	0.73	22.79	4.95	9	35	-0.98	(0.20)	.989
Year 5	602	0.17	23.05	4.93	9	35	1.78	(0.20)	.303
Year 6	458	1.08	23.25	4.89	0	35	-0.30	(0.22)	.989
GERT-S									
Year 1	575	0.52	29.14	4.18	18	38			
Year 2	446	0.67	29.49	4.20	17	38	1.44	(0.22)	.299
Year 3	581	0.34	30.24	4.04	18	40	3.43	(0.20)	**.002**
Year 4	545	0.55	30.81	4.29	17	40	2.89	(0.19)	**.012**
Year 5	603	0.00	31.45	3.93	17	41	3.54	(0.18)	**.002**
Year 6	460	0.65	31.34	4.37	17	42	1.05	(0.20)	.299

Significant p-values are indicated in bold. Comparisons to previous years were adjusted for multiple comparisons using the Holm-Bonferroni method

JSPE-S Jefferson Scale of Physician Empathy Student version, *QCAE* Questionnaire of Cognitive and Affective Empathy, *AMSP* Ability to Monitor Self-Presentation, *GERT-S* Geneva Emotion Recognition Test Short form

their medical school. Medical curricula have significantly changed in recent decades, and communication skills training has become a central part of medical education (see [42] for a detailed example). Our results indicate that improvement in empathy mainly occurs in the first three years of the six-year curriculum, suggesting that these first years are pivotal times for the development of interactional skills. At our university, a range of courses designed to foster empathy and enhance interaction skills in medical students have been integrated throughout the entire medical curriculum, including in the earlier years [42]. The improvement in empathy in the first years of medical school observed in our study might indicate the effect of such teaching. The results of the present study show a plateau during the last years of the curriculum, which corresponds to medical students' entry into clinical work [1]. Encountering the complexities of

clinical practice has previously been incriminated as the source of a drop in empathy. However, the present study's results indicate that medical students' empathy withstood these difficulties and remained stable in the second half of the curriculum. Additionally, the plateau observed in the present study may reflect a standard learning curve, where improvement in empathy is observed up to a maximum that cannot easily be surpassed for the time being.

Behavioral empathy was the only measure that showed stability over time. In the present study, the ability to modify one's behavior according to the social situation at hand (measured via the AMSP) was used as a proxy of behavioral empathy, which is usually defined as the ability to act accurately on the basis of one's understanding of others' emotions [11]. Given that it does not follow the same trajectory observed in the other empathy measures, this proxy might not totally measure the behavioral pendant of empathy and rather something related to self-monitoring abilities (as the AMSP instrument was originally created for). Any conclusion regarding behavioural empathy trajectories during medical school should be withheld until further studies explore other measures of this concept, the most pertinent being the observation of empathic behaviours during actual interactions.

The present study explored different psychosocial and health-related factors that could influence empathy trajectories, such as gender, parental education, social support, coping strategies, mental health, or physical activity. The results show that the influence of most of these factors on empathy trajectories was not substantial, which is surprising given that past studies have attested of the influence of several factors on empathy [14]. Gender identification was the only factor that had a substantial influence on the trajectories of empathy, with small effect sizes. We indeed observed that students identifying as male generally had lower trajectories of empathy (cognitive and affective) and emotion recognition than did students identifying as female. This gender difference in levels of empathy has been extensively reported in past studies and has been attributed to biological predispositions or sociocultural influences [43]. By accounting for a large array of potentially influential factors, the present study ensured a less biased estimation of empathy trajectories, but the most influential factors might be those not included: the factors pertaining to the educational environment. As a single-site study, the specific influence of the educational environment could not be analysed and is possibly the most determinant factor of medical students' empathy trajectories. Future multisite studies are needed to confirm this hypothesis.

Limitations

This study has several limitations that need to be mentioned. Although a more comprehensive approach to empathy is proposed by encompassing its cognitive, affective, and behavioral dimensions, as well as emotion recognition, the present study cannot claim to cover all aspects of this multifaceted and complex concept. Individuals with more socioeconomic and health issues are known to be less likely to participate in cohort studies [44] and might be underrepresented in our sample. Thus, despite a good response rate, empathy trajectories could differ among students who did not volunteer for the study. The longitudinal design, the analysis strategy (random-effects analysis), and the fact that multiple cohort (in terms of participation year) are represented at each analyzed time point have likely limited cohort effects, but these effects cannot be totally ruled out. The presence of a learning effect for the GERT-S cannot be completely dismissed. Nevertheless, learning effects for such a test from one year to the next are likely insignificant, and when looking solely at the students who participated once (and who thus could not have a learning effect), we could observe that students in more advanced curriculum years still had higher GERT-S scores than those in earlier curriculum years did.

Conclusion

The potential decrease in empathy during medical school has been largely broadcasted and raised awareness on the need for more interpersonal skills training. Since then, medical curricula have significantly evolved in that regard and might now be an environment more favourable to the development of interactional skills. Updated insights into the current status of medical students' empathy trajectories are thus needed. The present study indicates that empathy improves during medical school even when multiple dimensions of empathy are considered with different instruments, including an emotion recognition test that is not self-reported. Our results further warrant additional studies on the educational environment to identify the factors that most strongly influence medical students' empathy trajectories.

Abbreviations

AMSP	Ability to Modify Self-Presentation Scale
GERT-S	Geneva Emotion Recognition Test – Short Form
ICC	Intraclass correlation coefficient
IRI	Interpersonal Reactivity Index
JSPE	Jefferson Scale of Physician Empathy
JSPE-S	Jefferson Scale of Physician Empathy—Student Version
LMM	Linear Mixed Model
QCAE	Questionnaire of Cognitive and Affective Empathy
Std.	Standardized

Acknowledgements

The authors want to thank Sylvie Felix and Fabienne Thévenaz for their help in the ETMED-L project's data collection.

Authors' contributions
VC: data collection, data management, statistical analyses, and manuscript drafting; AB: project's principal investigator, project design, and manuscript drafting; CB: project design, data collection, and manuscript drafting; SR: statistical analyses and manuscript drafting; MB: establishment and refinement of project procedures, critical review of the manuscript; MSM: project design and critical review of the manuscript; SB, KS, JG, and PAB: establishment and refinement of project procedures and critical review of the manuscript. All the authors have read and approved the final manuscript.

Funding
Open access funding provided by University of Lausanne This work is supported by the Swiss National Science Foundation (grant number 10001C_197442). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability
The dataset supporting the conclusions of this article is available in the zenodo repository, http://doi.org/10.5281/zenodo.11277114.

Declarations

Ethics approval and consent to participate
The Cantonal Research Ethics Committee—Vaud (Commission cantonale d'éthique de la recherche sur l'être humain – Vaud [CER-VD]) approved the ETMED-L project (project number 2020–02474). The study was conducted in accordance with the Declaration of Helsinki principles and all participants provided written informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Psychiatric Liaison Service, Department of Psychiatry, Lausanne University Hospital (CHUV) and University of Lausanne, Avenue de Beaumont 23, 1011 Lausanne, Switzerland. [2]Service of General Psychiatry, Department of Psychiatry, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland. [3]Psychiatric Epidemiology and Psychopathology Research Centre, Department of Psychiatry, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland. [4]Institute of Psychology, University of Bern, Bern, Switzerland. [5]Addiction Medicine, Department of Psychiatry, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland. [6]Department of Internal Medicine, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland. [7]Department of Organizational Behavior, Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Lausanne, Switzerland.

Received: 14 August 2024 Accepted: 25 March 2025
Published online: 15 April 2025

References

1. Hojat M, Vergare MJ, Maxwell K, Brainard G, Herrine SK, Isenberg GA, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. Acad Med. 2009;84:1182–91.
2. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med. 2001;76:390–3.
3. Bachmann C, Abramovitch H, Barbu CG, Cavaco AM, Elorza RD, Haak R, et al. A European consensus on learning objectives for a core communication curriculum in health care professions. Patient Educ Couns. 2013;93:18–26.
4. Andersen FA, Johansen ASB, Søndergaard J, Andersen CM, Assing HE. Revisiting the trajectory of medical students' empathy, and impact of gender, specialty preferences and nationality: a systematic review. BMC Med Educ. 2020;20:52.
5. Ponnamperuma G, Yeo SP, Samarasekera DD. Is empathy change in medical school geo-socioculturally influenced? Med Educ. 2019;53:655–65.
6. Hojat M, Mangione S, Nasca TJ, Cohen MJM, Gonnella JS, Erdmann JB, et al. The jefferson scale of physician empathy: development and preliminary psychometric data. Educ Psychol Meas. 2001;61:349–65.
7. Spatoula V, Panagopoulou E, Montgomery A. Does empathy change during undergraduate medical education? – a meta-analysis. Med Teach. 2019;41:895–904.
8. Smith KE, Norman GJ, Decety J. The complexity of empathy during medical school training: Evidence for positive changes. Med Educ. 2017;51:1146–59.
9. Reniers RLEP, Corcoran R, Drake R, Shryane NM, Völlm BA. The QCAE: A Questionnaire of Cognitive and Affective Empathy. J Pers Assess. 2011;93:84–95.
10. Cuff BMP, Brown SJ, Taylor L, Howat DJ. Empathy: a review of the concept. Emot Rev. 2016;8:144–53.
11. Mercer SW, Reynolds WJ. Empathy and quality of care. Br J Gen Pract. 2002;52:S9-12.
12. Davis MH. A multidimensional approach to individual differences in empathy. Cat. Sel. Doc. Psychol. 1980;10.
13. Quince TA, Parker R, Wood DF, Benson JA. Stability of empathy among undergraduate medical students: a longitudinal study at one UK medical school. BMC Med Educ. 2011;11:90–8.
14. Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med. 2011;86:996–1009.
15. Quince TA, Thiemann P, Benson J, Hyde S. Undergraduate medical students' empathy: Current perspectives. Adv Med Educ Pract. 2016;7:443–55.
16. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. Med Educ. 2016;50:132–49.
17. Puthran R, Zhang MWB, Tam WW, Ho RC. Prevalence of depression amongst medical students: a meta-analysis. Med Educ. 2016;50:456–68.
18. Hope V, Henderson M. Medical student depression, anxiety and distress outside North America: a systematic review. Med Educ. 2014;48:963–79.
19. Carrard V, Bourquin C, Berney S, Schlegel K, Gaume J, Bart PA, et al. The relationship between medical students' empathy, mental health, and burnout: A cross-sectional study. Med Teach. 2022;44:1392–9.
20. Berney A, Carrard V, Berney S, Schlegel K, Gaume J, Gholam M, et al. Study protocol for the ETMED-L project: Longitudinal study of mental health and interpersonal competence of medical students in a Swiss university using a comprehensive framework of empathy. BMJ Open. 2021;11:e053070.
21. West BT, Welch KB, Galecki AT. Linear Mixed Models: A Practical Guide Using Statistical Software. 3rd ed. New York: Chapman and Hall/CRC; 2022.
22. Carrard V, Bourquin C, Berney A. Questionnaires for the ETMED-L project. 2024; https://doi.org/10.5281/zenodo.13361385
23. Myszkowski N, Brunet-Gouet E, Roux P, Robieux L, Malézieux A, Boujut E, et al. Is the Questionnaire of Cognitive and Affective Empathy measuring two or five dimensions? Evidence in a French sample. Psychiatry Res. 2017;255:292–6.
24. Myszkowski N, Storme M, Zenasni F, Lubart T. Appraising the duality of self-monitoring: Psychometric qualities of the Revised Self-Monitoring Scale and the Concern for Appropriateness Scale in French. Can J Behav Sci. 2014;46:387–96.
25. Lennox RD, Wolfe RN. Revision of the self-monitoring scale. J Pers Soc Psychol. 1984;46:1349–64.
26. Schlegel K, Grandjean D, Scherer KR. Introducing the Geneva Emotion Recognition Test: an example of Rasch-based test development. Psychol Assess. 2014;26:666–72.
27. Schlegel K, Boone RT, Hall JA. Individual differences in interpersonal accuracy: a Multi-level meta-analysis to assess whether judging other people is one skill or many. J Nonverbal Behav. 2017;41:103–37.
28 Voorpostel M, Tillmann R, Lebert F, Kuhn U, Lipps O, Ryser VA, et al. Swiss Household Panel Userguide (1999–2020), Wave 22. Lausanne: FORS; 2022.
29. Grob A, Bodmer NM, Flammer A. Living conditions in Europe: The case of Switzerland. Institute of Psychology, Bern: University of Bern; 1993.
30. Bodmer NM, Grob A. Well-being and constraints of adolescents: A comparison of adolescents from single-parent and two-parent families [Bien-être et contraintes d'adolescents: Une comparaison entre adolescents

de familles monoparentales et de familles biparentales]. Int J Psychol. 1996;31:39–48.
31. Perrin M, Vandeleur CL, Castelao E, Rothen S, Glaus J, Vollenweider P, et al. Determinants of the development of post-traumatic stress disorder, in the general population. Soc Psychiatry Psychiatr Epidemiol. 2014;49:447–57.
32. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Meas. 1977;1:385–401.
33. Spielberger CD. Manual for the State-trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press; 1983.
34. Schaufeli WB, Martínez IM, Pinto AM, Salanova M, Bakker AB. Burnout and engagement in university students: a cross-national study. J Cross-Cult Psychol. 2002;33:464–81.
35. Pinheiro J, Bates D. Mixed-Effects Models in S and S-PLUS. Springer Science & Business Media; 2006.
36. Gałecki A, Burzykowski T. Linear Mixed-Effects Model. In: Gałecki A, Burzykowski T, editors. Linear Mixed-Effects Models Using R: A Step-by-Step Approach. New York, NY: Springer; 2013. page 245–73.
37 Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
38. Lee KJ, Roberts G, Doyle LW, Anderson PJ, Carlin JB. Multiple imputation for missing data in a longitudinal cohort study: A tutorial based on a detailed case study involving imputation of missing outcome data. Int J Soc Res Methodol. 2016;19:575–91.
39. StataCorp. Stata Statistical Software: Release 17. StataCorp LLC. College Station, TX: 2021.
40. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: 2022.
41. Office fédéral de la statistique. Traitement pour problèmes psychiques en 2022. Off. Fédéral Stat.2023;Available from: https://www.bfs.admin.ch/asset/fr/28465393. Cited 2023 Dec 5
42. Berney A, Bourquin C. Individual supervision to enhance reflexivity and the practice of patient-centered care: experience at the undergraduate level. J Cancer Educ. 2019;34:363–5.
43. Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: Gender effects in brain and behavior. Neurosci Biobehav Rev. 2014;46:604–27.
44. Rothenbühler M, Voorpostel M. Attrition in the Swiss Household Panel: Are Vulnerable Groups more Affected than Others?. In: Oris M, Roberts C, Joye D, Ernst Stähli M, editors. Surveying Human Vulnerabilities across the Life Course. Cham: Springer International Publishing; 2016. pp. 223–44.: https://doi.org/10.1007/978-3-319-24157-9_10 Cited 2024 Jan 9

Publisher's Note