Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Professionalism in the Gray Zone: Comparing Medical Student and Preceptor Judgments of Ambiguous Clinical Scenarios

Terry L. Hudgins

thudgins@rvu.edu

Rocky Vista University

Elisabeth Kac

Rocky Vista University College of Osteopathic Medicine

C. Ping Huang

Rocky Vista University

Andrea Haan

Rocky Vista University

Research Article

Keywords: Medical professionalism, undergraduate medical education, medical students, clinical preceptors, professionalism education, professional identity formation, hidden curriculum, clinical vignettes

Posted Date: October 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10941126/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Professionalism is a core competency in medical education, yet many challenges encountered during clinical training occur in ambiguous situations where expectations depend on clinical context, role responsibilities, and workplace norms. These “gray-zone” situations may be particularly challenging for students transitioning into clerkships. This study examined whether medical students and physician clinical preceptors differed in their judgments of professionalism across gray-zone clinical scenarios and in their confidence in those judgments.

Methods: A cross-sectional survey was administered to second-year osteopathic medical students and physician clinical preceptors. Participants classified behaviors depicted in 20 standardized professionalism vignettes as professional or unprofessional and rated their confidence in each judgment. Responses from 75 students and 185 preceptors (N = 260) were analyzed. Vignette-level classifications were compared using Pearson’s chi-square or Fisher’s exact tests with Benjamini-Hochberg false discovery rate correction. Domain-level misalignment was summarized using mean absolute percentage-point differences. Participant-level mean confidence scores were also compared between groups.

Results: Seven of 20 vignettes demonstrated student–preceptor differences at the unadjusted significance threshold, and four remained significant after correction for multiple comparisons. The largest difference involved accessing a patient’s medical record after direct clinical involvement had ended, classified as professional by 10.7% of students compared with 61.6% of preceptors. Differences also remained for follow-up of imaging results, phone use during rounds, and use of the term “noncompliant” without contextual framing. The greatest domain-level misalignment occurred among vignettes involving integrity, honesty, and trustworthiness (20.0 percentage points), with additional differences involving patient-centered language, scope and role clarity, and accountability. Students demonstrated lower overall confidence than preceptors (2.43± 0.24 vs. 2.62 ± 0.27; Hedges’ g = 0.72; $p < .000001$); however, confidence differences did not consistently parallel differences in professionalism judgments.

Conclusions: Medical students and physician preceptors frequently agreed about professionalism but differed substantially in selected ambiguous clinical situations. These differences were not consistently attributable to lower student confidence, suggesting that students and preceptors may apply different contextual assumptions when interpreting the same behavior. Explicitly addressing gray-zone scenarios and the reasoning underlying different professional judgments may help make implicit clinical expectations more visible and better prepare students to navigate context-dependent professionalism during clinical training.

Background

Medical professionalism is foundational to clinical education, safe patient care, and trust among patients, physicians, and health care teams. It encompasses accountability, integrity, communication,

respect, and patient-centered care and is a core competency in determining readiness for independent practice [1–3]. Yet translating professionalism into observable behaviors that can be taught and assessed consistently remains challenging [3].

Professional expectations are not uniform across clinical environments. Specialty, service, supervising clinician, workplace culture, and local norms can influence how the same behavior is interpreted [4]. Students must therefore navigate expectations involving communication, teamwork, hierarchy, accountability, professional boundaries, and patient advocacy. When expectations are implicit, ambiguity may reduce psychological safety and make learners less likely to seek clarification or voice concerns [5]. Instead, students may rely on observation, peer behavior, role modeling, and the hidden curriculum—the informal lessons conveyed through workplace culture and interactions—to determine what constitutes professional conduct [6, 7]. These experiences, along with pre-existing personal and entering professional identities, contribute to professional identity development alongside formal instruction [8, 9].

This variability in professional expectations aligns with a broader shift in medical education from viewing professionalism primarily as a set of personal attributes and observable behaviors to understanding it as part of professional identity formation. Professional identity develops through socialization, role modeling, relationships, experience, and reflection, reinforcing the importance of the clinical learning environment in shaping how learners understand and enact professionalism [10]. Contemporary frameworks emphasize the developmental process through which learners internalize professional values, norms, and responsibilities and apply them in clinical practice [10–12]. Professional judgment is one expression of this development, reflecting how learners interpret and prioritize professional values in particular situations. Accordingly, professionalism is increasingly understood not simply as a checklist of behaviors but as the capacity to exercise sound judgment within complex clinical environments.

Recent literature indicates that judgments of professionalism are influenced not only by the behavior itself, but also by clinical context, competing professional priorities, cultural and institutional norms, and the perspectives and biases of those interpreting the behavior [13–16]. Such differences create challenges for assessment and feedback because identical behaviors may be interpreted differently across supervisors or clinical settings. Learners may consequently encounter conflicting expectations during clinical training. This issue is particularly important when concerns arise in ambiguous situations, where students may hesitate to raise questions because of perceived consequences for relationships, team dynamics, or evaluation [17].

We use the term “gray-zone professionalism” to describe situations in which the interpretation of professional behavior depends on clinical context, implicit workplace expectations, and individual judgment. Unlike clear violations, gray-zone situations may involve incomplete information, competing professional values, or more than one defensible interpretation. Disagreement in such cases may therefore reflect differences in contextual reasoning rather than deficiencies in professionalism knowledge. Examples include determining when to escalate a concern, how assertively to communicate

within a hierarchy, how to balance efficiency with patient-centered communication, how to use technology during patient care, and how to navigate professional boundaries. These types of ambiguous and context-dependent situations are a recognized feature of professionalism in clinical education, where students and faculty may differ in how they interpret expectations and what constitutes appropriate professional behavior [18, 19].

Despite increasing recognition of the contextual nature of professionalism, limited research has directly examined whether medical students and clinical preceptors interpret the same ambiguous situations similarly. Direct comparison is important because differences in interpretation may contribute to inconsistent feedback and assessment and influence the development of professional judgment. Standardized vignettes provide an opportunity to examine these differences by presenting participants with identical scenarios while preserving the ambiguity characteristic of real clinical encounters [20].

This study addresses this underexplored area by directly comparing medical students' and physician clinical preceptors' judgments of professionalism across standardized gray-zone clinical vignettes. By presenting both groups with identical scenarios, we examine areas of agreement and disagreement and compare confidence in those judgments. This approach allows us to examine whether differences in interpretation emerge when students and preceptors are asked to evaluate the same ambiguous situations, rather than inferring differences from separate assessments or experiences. Understanding where perspectives converge and diverge may clarify how contextual assumptions, role expectations, and professional values influence judgments of professionalism and may help explain inconsistencies in feedback and assessment. Rather than assuming that one group holds the "correct" interpretation, examining these differences can make implicit expectations visible and create opportunities for dialogue about the professional principles, contextual factors, and role responsibilities that shape judgment. These insights may inform educational approaches that better prepare learners to recognize, discuss, and navigate ambiguity during clinical training.

Methods

This cross-sectional survey study compared judgments of professionalism between second-year osteopathic medical students and physician clinical preceptors using standardized clinical vignettes depicting common "gray-zone" professionalism scenarios. The study was approved by the Rocky Vista University Institutional Review Board (IRB #2026-059).

Participants and Setting

Participants included second-year (OMS II) medical students enrolled at Rocky Vista University College of Osteopathic Medicine and physician preceptors who supervise third- and fourth-year osteopathic medical students during clinical clerkships.

Student participants were recruited through institutional email announcements distributed to the Class of 2028. Physician preceptors representing multiple core clinical specialties were recruited through

departmental and clerkship email distribution lists. Participation was voluntary, and no compensation was provided.

Eligible physician participants included actively practicing clinicians serving as clinical preceptors in one of the core clerkship specialties, including Family Medicine, Internal Medicine, Obstetrics and Gynecology, Pediatrics, Psychiatry, Surgery, or another clinical specialty.

Survey Instrument

The survey instrument was developed specifically for this study and has not been previously published. The survey included demographic questions and 20 vignette-based scenarios depicting potentially ambiguous (“gray-zone”) professionalism situations encountered in undergraduate medical education. Scenarios represented multiple professionalism domains, including communication and teamwork; accountability and responsibility; integrity, honesty, and trustworthiness; respect, boundaries, and team culture; patient-centeredness and contextual language; self-awareness and professional development; and scope and role clarity. Participants classified each behavior as Professional or Unprofessional and rated their confidence as Very Confident, Somewhat Confident, or Not Confident. The dichotomous classification was intentionally used to require a categorical professionalism judgment rather than an assessment of severity, while confidence ratings assessed participants’ certainty in those judgments.

Vignette Development

The initial 20 clinical vignettes were drafted with AI assistance using investigator-developed prompts reflecting professionalism scenarios encountered in undergraduate medical education. AI was used only to assist with initial narrative development and improve readability and consistency. The investigative team developed the vignette concepts and professionalism constructs, substantially revised all content, assigned each scenario to a primary professionalism domain by consensus, and retained responsibility for all scientific and methodological decisions.

Content validation was conducted by five practicing physician preceptors representing core clinical clerkship disciplines. Reviewers independently assessed each vignette for clinical realism, appropriateness for medical students, focus on professionalism rather than clinical competence, suitability for inclusion, and whether reasonable clinicians could disagree about the behavior's classification. Two relatively unambiguous scenarios were intentionally included as benchmark items to provide an internal check that participants interpreted the classification task as intended; these were not considered formal psychometric validation.

Vignettes were revised iteratively based on reviewer feedback to improve clarity and authenticity while preserving the intended ambiguity. The final study-developed instrument consisted of 20 standardized scenarios representing multiple professionalism domains encountered during clinical education. The complete English-language survey instrument is provided as Additional file 1.

Data Collection

The anonymous survey was administered electronically through Qualtrics following electronic informed consent. No identifying information was collected. Demographic variables included participant role, gender identity, age category, and physician specialty for preceptors. Data were exported from Qualtrics to a secure, password-protected Rocky Vista University Microsoft Teams server accessible only to study investigators.

An a priori power analysis using G*Power (Version 3.1.9.7; Heinrich Heine University, Düsseldorf, Germany) indicated that 88 participants (44 per group) were required to detect a medium effect (Cramer's $V = 0.30$) using a chi-square test of independence with $\alpha = .05$ and 80% power. All eligible OMS II students and physician preceptors were invited to participate to maximize precision and permit exploratory subgroup analyses. The final sample included 75 students and 185 physician preceptors (N = 260), exceeding the minimum required sample.

Statistical Analysis

Descriptive statistics summarized participant characteristics and response frequencies. At the vignette level, proportions of students and physician preceptors classifying each behavior as professional were compared using Pearson's chi-square test or Fisher's exact test when expected cell counts were fewer than five. Odds ratios (ORs) and 95% confidence intervals (CIs) quantified group differences, with ORs representing the odds of students classifying a behavior as professional relative to preceptors. Statistical significance was initially set at $p < .05$, with the Benjamini-Hochberg false discovery rate (FDR) procedure applied across the 20 vignette-level comparisons.

The 20 vignettes were also grouped into predefined professionalism domains based on their primary constructs. Domain-level student–preceptor misalignment was summarized descriptively as the mean absolute percentage-point difference in professionalism classifications across vignettes within each domain. Larger values indicated greater divergence between groups. Domain-level comparisons were descriptive; inferential statistical testing was not performed.

Confidence responses were coded as 1 = Not Confident, 2 = Somewhat Confident, and 3 = Very Confident. A mean confidence score across the 20 vignettes was calculated for each participant with complete responses and compared between groups using Welch's independent-samples t test. Mean differences with 95% CIs and Hedges' g were reported. A Mann–Whitney U test was conducted as a sensitivity analysis given the ordinal response scale. Confidence distributions for vignettes that remained significant after FDR correction were also examined descriptively.

Results

A total of 260 participants were included in the analysis, consisting of 75 second-year osteopathic medical students and 185 physician clinical preceptors. Among students, 46 (61.3%) identified as female

and 29 (38.7%) as male. Among preceptors, 68 (36.8%) identified as female, 113 (61.1%) as male, and 4 (2.2%) preferred not to answer. Most students were between 18 and 26 years of age (57.3%), whereas most preceptors were between 43 and 58 years of age (53.0%). Preceptors represented multiple clinical specialties and practice settings and had substantial experience in both clinical practice and medical student education (Tables 1 and 2).

Table 1
Demographic Characteristics of Study Participants, Stratified by Respondent Group

Variable	Total		Students		Preceptors	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Gender Identity						
Male	142	54.6	29	38.7	113	61.1
Female	114	43.8	46	61.3	68	36.8
Other/Prefer not to answer	4	1.5	0	0.0	4	2.2
Age						
18–26 years	43	16.5	43	57.3	0	0.0
27–42 years	73	28.1	31	41.3	42	22.7
43–58 years	99	38.1	1	1.3	98	53.0
59–77 years	40	15.4	0	0.0	40	21.6
78 + years	5	1.9	0	0.0	5	2.7

Preceptors represented a range of clinical specialties and practice settings. Family medicine was the most frequently represented specialty (n = 55, 29.7%), followed by surgery (n = 24, 13.0%) and internal medicine (n = 23, 12.4%). The most common practice settings were private practice (n = 62, 33.5%) and community hospitals (n = 61, 33.0%). Preceptor specialty and practice setting characteristics are presented in Table 2.

Table 2
Clinical Specialty and Practice Setting of Preceptors Included in the Study

Variable	Frequency	Percent
Specialty		
Family Medicine	55	29.7
Internal Medicine	23	12.4
Pediatrics	17	9.2
Ob/Gyn	11	5.9
Surgery	24	13.0
Psychiatry	19	10.3
Other	36	19.5
Practice Setting		
Academic Center	19	10.3
Community Hospital	61	33.0
Private Practice	62	33.5
Rural Setting	16	8.6
VA/Military	6	3.2
Other	21	11.4

Professionalism classifications differed between medical students and clinical preceptors across several clinical scenarios (Table 3). At the unadjusted significance threshold ($p < .05$), seven of the 20 vignettes demonstrated statistically significant differences between groups (Vignettes 3, 4, 5, 7, 8, 12, and 17). After Benjamini-Hochberg false discovery rate correction for multiple comparisons, four differences remained statistically significant (Vignettes 3, 5, 7, and 17). Responses to the two intentionally included benchmark vignettes (Vignettes 2 and 13) demonstrated high agreement with their intended classifications, providing an internal check that participants interpreted the professionalism classification task as intended.

The largest difference was observed for the scenario involving chart access after direct patient care had ended (Vignette 5), which was classified as professional by 10.7% of students compared with 61.6% of preceptors ($p < .000001$). Differences were also observed for failure to follow up on imaging results (Vignette 3), with 20.0% of students and 47.0% of preceptors classifying the behavior as professional ($p = .00005$); phone use during rounds (Vignette 7), which was classified as professional by 41.3% of students compared with 60.5% of preceptors ($p = .0048$); and use of the term “noncompliant” without

contextual framing (Vignette 17), which was classified as professional by 10.7% of students and 28.1% of preceptors (p = .0025). These four differences remained statistically significant after adjustment for multiple comparisons.

Three additional scenarios demonstrated significant differences between students and preceptors prior to adjustment for multiple comparisons. Copying portions of a prior physical examination into a progress note without rechecking every documented system (Vignette 4) was classified as professional by 0.0% of students compared with 5.9% of preceptors (p = .0370). Differences were also observed for use of casual language and humor with the resident team during rounds (Vignette 8), which was classified as professional by 33.3% of students compared with 48.1% of preceptors (p = .0296), and continuing the same oral presentation style after receiving feedback that presentations were too detailed (Vignette 12), which was classified as professional by 4.0% of students and 13.5% of preceptors (p = .0242). These differences did not remain statistically significant after adjustment for multiple comparisons.

Table 3
Student and Preceptor Professionalism Classifications by Vignette

Vignette	Student Professionalism %	Preceptor Professionalism %	OR (95% CI)	p-value	FDR Adjusted p
1	18.7	21.1	0.82 (0.41–1.63)	0.6615	.662
2	96.0	99.5	0.12 (0.01–1.20)	0.0737	.164
3	20.0	47.0	0.27 (0.14–0.53)	0.00005	< .001
4	0.0	5.9	0.11 (0.01–1.83)	0.0370	.106
5	10.7	61.6	0.08 (0.04–0.17)	< 0.000001	< .001
6	9.3	18.4	0.45 (0.19–1.05)	0.0698	.164
7	41.3	60.5	0.46 (0.27–0.80)	0.0048	.024
8	33.3	48.1	0.54 (0.31–0.93)	0.0296	.099
9	81.3	85.4	0.74 (0.35–1.57)	0.4144	.513
10	26.7	35.7	0.66 (0.37–1.18)	0.1619	.270
11	5.3	8.1	0.64 (0.20–2.08)	0.4361	.513
12	4.0	13.5	0.27 (0.08–0.89)	0.0242	.097
13	98.7	99.5	0.20 (0.01–3.30)	0.4945	.549

Vignette	Student Professionalism %	Preceptor Professionalism %	OR (95% CI)	p-value	FDR Adjusted p
14	5.3	2.2	2.56 (0.48–13.70)	0.2332	.333
15	90.7	93.5	0.68 (0.25–1.82)	0.4243	.513
16	66.7	74.6	0.68 (0.38–1.23)	0.1515	.270
17	10.7	28.1	0.31 (0.14–0.66)	0.0025	.0166
18	58.7	54.6	1.18 (0.68–2.04)	0.5492	.578
19	1.3	6.5	0.19 (0.02–1.51)	0.1167	.233
20	0.0	3.2	0.38 (0.02–7.88)	0.1862	.286

Differences in professionalism classifications between students and preceptors varied across professionalism domains (Table 4). The greatest mean absolute difference was observed for Integrity, Honesty, and Trustworthiness (20.0 percentage points), followed by Patient-Centeredness and Contextual Language (12.7 percentage points), Scope and Role Clarity (11.2 percentage points), and Accountability and Responsibility (10.9 percentage points). Smaller differences were observed for Respect, Boundaries and Team Culture (9.3 percentage points), Self-Awareness and Professional Development (4.8 percentage points), and Communication and Teamwork (3.7 percentage points).

Table 4
Domain-Level Professionalism Misalignment

Domain	Vignettes	Individual Absolute Differences	Mean Absolute Differences
Integrity, Honesty, and Trustworthiness	4, 5, 14	5.9, 50.9, 3.1	20.0 pp
Patient-Centeredness and Contextual Language	16, 17	8.7 17.4	13.1 pp
Scope and Role Clarity	7, 20	19.2, 3.2	11.2 pp
Accountability and Responsibility	3, 11, 15	27.0, 2.8, 2.8	10.9 pp
Respect, Boundaries, and Team Culture	8, 9, 10	14.8, 4.1, 9.90	9.3 pp
Self-Awareness and Professional Development	12, 13, 18	9.6, 0.8, 4.1	4.8 pp
Communication and Teamwork	1, 2, 19	2.4, 3.5, 5.2	3.7 pp

Note. Domain-level misalignment was calculated as the mean of the absolute percentage-point differences between student and preceptor professionalism classifications for the vignettes assigned to each domain. Larger values indicate greater divergence between student and preceptor judgments. These values are descriptive and were not subjected to inferential significance testing.

Students demonstrated lower overall confidence than physician preceptors when evaluating the professionalism scenarios (2.43 ± 0.24 vs. 2.62 ± 0.27, respectively). The mean difference was 0.19 points (95% CI, 0.12–0.26), representing a moderate-to-large standardized difference (Hedges' $g = 0.72$; $p < .000001$). The Mann–Whitney U sensitivity analysis yielded a consistent between-group difference ($p < .000001$). Overall confidence results are presented in Table 5.

Table 5
Confidence in Professionalism Classifications by Respondent Group

Measure	Students	Preceptors
Participants with complete confidence data	75	185
Overall confidence score, mean ± SD	2.43 ± 0.24	2.62 ± 0.27
Median confidence score (IQR)	2.45 (2.28–2.60)	2.63 (2.45–2.85)
Very Confident, % of complete ratings	51.8%	66.9%
Somewhat Confident, % of complete ratings	39.5%	28.2%
Not Confident, % of complete ratings	8.7%	4.9%

Note. Confidence was coded as 1 = Not Confident, 2 = Somewhat Confident, and 3 = Very Confident. Participant-level mean confidence scores were calculated across the 20 vignettes. The mean difference between groups was 0.19 points (95% CI, 0.12–0.26), with greater confidence among physician preceptors (Hedges' g = 0.72; $p < .000001$).

Differences in confidence varied across the four scenarios that demonstrated significant student–preceptor differences in professionalism classification after FDR correction. For the scenario involving failure to follow up on imaging results (Vignette 3), 38.7% of students compared with 67.0% of preceptors reported being very confident. Similarly, for phone use during rounds (Vignette 7), 22.7% of students compared with 48.6% of preceptors reported being very confident. In contrast, the proportion reporting being very confident was nearly identical between students and preceptors for chart access after direct patient care had ended (Vignette 5; 58.7% vs. 58.9%, respectively). For use of the term “noncompliant” without contextual framing (Vignette 17), 56.0% of students compared with 50.3% of preceptors reported being very confident

Discussion

This study found that medical students and physician preceptors generally shared broad expectations of professionalism yet diverged in selected gray-zone situations in which intent, role responsibility, workflow, or clinical context could reasonably shape interpretation. Misalignment was greatest in scenarios involving integrity and trustworthiness, with additional differences involving accountability, role clarity and technology use, and patient-centered language, while communication and teamwork showed comparatively greater agreement. Preceptors also reported greater confidence overall, but confidence did not consistently track with disagreement, suggesting that misalignment cannot be explained simply by learners being less certain. Students and experienced clinicians may instead apply different contextual assumptions while endorsing similar overarching professional values.

Because participants were not asked to explain the reasoning behind their classifications, the mechanisms underlying these differences cannot be established from the present data. One plausible interpretation is that students earlier in training rely more heavily on explicit rules, whereas experienced

clinicians draw more readily on contextualized judgment developed through repeated participation in clinical work. Preclinical professionalism curricula appropriately emphasize recognizable expectations such as confidentiality, respectful communication, responsibility, and boundaries. These foundations remain essential, but clinical practice often requires deciding how they apply when responsibilities overlap, information is incomplete, or values compete. The observed differences may therefore reflect variation in how respondents translated general standards into judgments about particular situations rather than disagreement about the standards themselves.

This interpretation is consistent with contemporary models of professional identity formation. Professionalism scholarship has moved beyond viewing professionalism solely as a set of observable behaviors toward understanding how learners internalize professional values, norms, and responsibilities through socialization, experience, relationships, and reflection [10–12]. More recent work further emphasizes that professional identity develops through active engagement with, rather than passive adoption of, professional norms [9]. From this perspective, differences between second-year students and practicing preceptors are developmentally plausible. Students approaching clerkships have had fewer opportunities to interpret professional principles alongside team structure, clinical responsibility, patient needs, or local workflow, whereas preceptors have accumulated experience doing so. Contextualized professional judgment may therefore be one observable feature of professional identity development.

The findings also align with literature on the hidden curriculum. Formal instruction communicates stated expectations, whereas the clinical learning environment transmits additional messages through role modeling, workplace routines, hierarchy, and locally negotiated norms [6, 7]. Students may consequently encounter different interpretations of appropriate behavior across services or supervisors. Stubbing et al. [4] described tensions learners experience when navigating professionalism expectations, Doherty et al. [5] demonstrated how students learn to manage unprofessional behavior through the clinical environment, and Poola et al. [18] showed that perceptions of professionalism vary by setting. The current study complements these findings by showing that variation may be evident even when students and preceptors evaluate identical written scenarios. Thus, inconsistency may reflect both different clinical norms and different interpretive frameworks.

The vignette methodology also connects this study to emerging work comparing learner and educator perceptions using standardized scenarios. Colón-López and Zinski [20], for example, found that students and preceptors did not always agree when judging learner mistreatment across identical vignettes. Although mistreatment and professionalism are distinct constructs, both studies show the value of identical scenarios for revealing different assumptions and thresholds. Goodwin and Oliver [19] likewise reported differences between student and faculty expectations when benchmarking adequate professionalism across stages of undergraduate medical education. Together, these studies suggest that professionalism expectations are not uniformly interpreted and that disagreement can reveal otherwise implicit assumptions.

The current results also reinforce longstanding challenges in professionalism assessment. Professionalism judgments can be shaped by context, observer interpretation, institutional culture, and the behaviors selected for evaluation [3, 21, 22]. When different observers interpret the same behavior differently, variability in ratings may reflect differences among evaluators as well as differences among learners. A purely categorical label of professional or unprofessional can therefore obscure information about intent, severity, role expectations, consequences, and whether expectations had previously been communicated. The binary format exposed thresholds of disagreement but also illustrates why authentic assessment may require richer contextual description and narrative explanation.

Educationally, these findings support preparing students for clerkships through explicit discussion of ambiguous cases rather than relying primarily on lists of acceptable and unacceptable behaviors. Gray-zone cases can invite learners to identify relevant principles, surface assumptions, consider what information would change their judgment, and distinguish institutional requirements from specialty norms or individual preferences. The goal should not be to teach that the preceptor response is inherently the correct one. Instead, discussion can help students recognize when a situation is context dependent and when clarification is appropriate, while maintaining non-negotiable ethical and professional standards.

This approach informed the Ask-Align-Act framework developed for the next phase of the project. Ask encourages learners to recognize uncertainty and clarify expectations when appropriate; Align prompts them to consider the clinical context, team expectations, and applicable professional standards; and Act emphasizes selecting a professionally defensible response based on that information. The framework is intended as a reasoning scaffold rather than a prescriptive answer key. It may be particularly useful during the transition to clerkships, when learners encounter new hierarchies and workflows and may not know which expectations are universal, institutionally mandated, or locally preferred. Structured case discussion could make implicit expectations more visible and give students language for seeking clarification.

Faculty development is an equally important implication. Complete agreement among preceptors is neither realistic nor necessarily desirable when genuine contextual ambiguity exists, but faculty can be more explicit about the basis for their judgments. Calibration could distinguish core ethical or institutional standards from specialty conventions, team practices, and individual preferences. Feedback should likewise describe behavior, context, and impact rather than relying on the global label “unprofessional,” reducing the risk that a supervisor's preference is interpreted as a universal rule. It may also support psychological safety by normalizing appropriate clarification when expectations are uncertain, an issue that is especially relevant given evidence that learners may hesitate to raise concerns because of hierarchy or perceived consequences [17].

Several strengths support the contribution of this study. Students and preceptors evaluated the same standardized scenarios, reducing case variation as an explanation for group differences. The vignettes intentionally represented ambiguous situations rather than obvious violations and underwent structured

review for realism, learner appropriateness, and genuine potential for disagreement. Examining vignette- and domain-level patterns identified areas useful for curricular discussion, while confidence ratings showed that disagreement was not consistently synonymous with uncertainty.

The findings should also be interpreted in light of important limitations. This was a single-institution, cross-sectional study, so differences cannot be attributed specifically to clinical experience or professional identity development, and local culture may limit generalizability. Differences in age, professional role, and accumulated experience could also contribute to judgments. Voluntary participation introduces potential response bias, and preceptor representation was not evenly distributed across specialties. The categorical response format simplified complex judgments and did not capture participants' reasoning, degrees of concern, or the contextual assumptions they made. Finally, standardized vignettes necessarily omit some of the relational and organizational detail available in real clinical encounters.

Future work should examine why participants reach different judgments and whether those judgments change as students progress through clinical training. Qualitative interviews, think-aloud methods, or open-ended responses could identify which professional values, contextual cues, and role assumptions drive disagreement. Longitudinal studies could test whether students become more context sensitive during clerkships without assuming that increasing similarity to preceptors is itself the desired endpoint. Multi-institutional studies could further distinguish shared expectations from local norms. The next phase will evaluate a structured gray-zone professionalism curriculum using Ask-Align-Act, examining reasoning, confidence, willingness to clarify expectations, and learners' experiences in clinical settings.

Conclusion

This study identified meaningful differences in how medical students and physician preceptors judged professionalism in ambiguous clinical situations. The greatest student–preceptor misalignment occurred among vignettes involving integrity, honesty, and trustworthiness, followed by patient-centeredness and contextual language, scope and role clarity, and accountability and responsibility, while communication and teamwork showed comparatively little disagreement. Although preceptors demonstrated greater overall confidence, differences in confidence did not consistently parallel differences in professionalism judgments, suggesting that misalignment cannot be explained solely by greater uncertainty among students. Rather, students and preceptors may share broad professional principles yet differ in how they apply them when expectations depend on clinical context, roles, workplace norms, and individual interpretation. Neither group’s responses should be considered inherently “correct”; instead, these differences highlight the limitations of relying on implicit expectations or the hidden curriculum to prepare students for clinical training. Making these expectations more explicit and providing learners with structured opportunities to examine ambiguous scenarios may strengthen contextual professional reasoning and better prepare students to navigate gray-zone professionalism during clerkships.

Abbreviations

AI
artificial intelligence
CI
confidence interval
FDR
false discovery rate
IQR
interquartile range
IRB
Institutional Review Board
OMS II
second-year osteopathic medical student
OR
odds ratio
SD
standard deviation
VA
Veterans Affairs.

Declarations

Human Ethics and Consent to Participate Declaration

This study was reviewed and approved by the Rocky Vista University Institutional Review Board (IRB #2026-059). All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki and applicable institutional guidelines and regulations. Participation was voluntary, and all participants provided informed consent prior to participation.

Ethics approval and consent to participate

This study was reviewed and approved by the Rocky Vista University Institutional Review Board (IRB #2026-059). All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki and applicable institutional guidelines and regulations. Participation was voluntary, and all participants provided informed consent prior to participation.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

TH: Conceptualization, Methodology, Investigation, Writing – Review & Editing. EK: Writing – Original Draft, Writing – Review & Editing. CPH: Writing – Review & Editing. AH: Writing – Original Draft, Writing – Review & Editing.

Acknowledgements

Not applicable.

Authors' information

Not applicable.

References

1. Desai MK, Kapadia JD. Medical professionalism and ethics. J Pharmacol Pharmacother. 2022;13(2):113–118. doi:10.1177/0976500X221111448.
2. Edgar L, Hatlak K, Haynes IL, Holmboe ES, Hogan SO, McLean S. The milestones guidebook. Accreditation Council for Graduate Medical Education; 2025.
3. Jaramillo Villegas C, Hernandez Rincon EH, Gallego Beltran JC, Guzman Sabogal YR. Assessment of professionalism in undergraduate and graduate medical education: a scoping review. BMC Med Educ. 2025;25:1295. doi:10.1186/s12909-025-07756-w.
4. Stubbing EA, Helmich E, Cleland J. Medical student views of and responses to expectations of professionalism. Med Educ. 2019;53(10):1025–1036. doi:10.1111/medu.13933.
5. Doherty A, Urwin R, McMullan RD, Tou YY, Westbrook JI, Churruca K. The hidden curriculum in which medical students learn to understand and manage unprofessional behaviour: a qualitative interview study. Med Sci Educ. 2025;35(1):415–423. doi:10.1007/s40670-024-02208-4.
6. Hafferty FW, Gaufberg EH, O'Donnell JF. The role of the hidden curriculum in "On Doctoring" courses. AMA J Ethics. 2015;17(2):129–137. doi:10.1001/virtualmentor.2015.17.2.medu1-1502.

7. Lehmann LS, Sulmasy LS, Desai S. Hidden curricula, ethics, and professionalism: optimizing clinical learning environments in becoming and being a physician. Ann Intern Med. 2018;168(7):506–508. doi:10.7326/M17-2058.
8. Lu V, Kumar K. The hidden curriculum of peer teaching in developing a professional identity: perspectives of medical students and junior doctors. Clin Teach. 2024;21(2):e13680. doi:10.1111/tct.13680.
9. Sternszus R, Steinert Y, Razack S, Boudreau JD, Snell L, Cruess RL. Being, becoming, and belonging: reconceptualizing professional identity formation in medicine. Front Med (Lausanne). 2024;11:1438082. doi:10.3389/fmed.2024.1438082.
10. Cruess SR, Cruess RL, Steinert Y. Supporting the development of a professional identity: general principles. Med Teach. 2019;41(6):641–649. doi:10.1080/0142159X.2018.1536260.
11. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. Reframing medical education to support professional identity formation. Acad Med. 2014;89(11):1446–1451. doi:10.1097/ACM.0000000000000427.
12. Irby DM, Hamstra SJ. Parting the clouds: three professionalism frameworks in medical education. Acad Med. 2016;91(12):1606–1611. doi:10.1097/ACM.0000000000001190.
13. Maristany D, Hauer KE, Leep Hunderfund AN, Elks ML, Bullock JL, Kumbamu A, et al. The problem and power of professionalism: a critical analysis of medical students' and residents' perspectives and experiences of professionalism. Acad Med. 2023;98(Suppl 3):S32–S41. doi:10.1097/ACM.0000000000005367.
14. Kassam A, de Vries I, Zabar S, Durning SJ, Holmboe E, Hodges B, et al. The next era of assessment within medical education: exploring intersections of context and implementation. Perspect Med Educ. 2024;13(1):496–506. doi:10.5334/pme.1128.
15. Shah D, Behravan N, Al-Jabouri N, Sibbald M. Incorporating equity, diversity and inclusion (EDI) into the education and assessment of professionalism for healthcare professionals and trainees: a scoping review. BMC Med Educ. 2024;24:991. doi:10.1186/s12909-024-05981-3.
16. Song X, Willy MJ. Exploring unprofessional behaviors and biased perceptions in the clinical environment: students' perspectives. Med Sci Educ. 2024;34(5):1049–1057. doi:10.1007/s40670-024-02087-9.
17. Sullivan E, Thampy H, Gay S. Raising professionalism concerns as a medical student: damned if they do, damned if they don't? BMC Med Educ. 2024;24:208. doi:10.1186/s12909-024-05144-4.
18. Poola VP, Suh B, Parr T, Boehler M, Han H, Mellinger J. Medical students' reflections on surgical educators' professionalism: contextual nuances in the hidden curriculum. Am J Surg. 2021;221(2):270–276. doi:10.1016/j.amjsurg.2020.09.003.
19. Goodwin AM, Oliver SW. Benchmarking "adequate professionalism" during undergraduate medical education: an exploratory survey study. J Med Educ Curric Dev. 2026;13:23821205251408682. doi:10.1177/23821205251408682.

20. Colón-López A, Zinski A. In which scenarios do preceptors and students agree, disagree, or remain neutral about learner mistreatment? Med Sci Educ. 2025;35:949–958. doi:10.1007/s40670-024-02261-z.
21. Frohna JG, Padmore JS. Assessment of professionalism in the graduate medical education environment. J Grad Med Educ. 2021;13(2 Suppl):81–85. doi:10.4300/JGME-D-20-00845.1.
22. Goldie J. Assessment of professionalism: a consolidation of current thinking. Med Teach. 2013;35(2):e952–e956. doi:10.3109/0142159X.2012.714888.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ProfessionalismintheGreyZoneSurvey.docx
- ProfessionalismintheGreyZone.docx