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.

Evaluative (Mis)alignment in Competency-Based Feedback: A Constructivist Grounded Theory-Informed Study in Family and Community Medicine Residency Programs

Hamana Oliveira Queiroz Bessa
Universidade Estadual da Paraíba
Wellington Pedro de Sousa
Universidade Estadual da Paraíba
Edinete Nunes da Silva
Universidade Estadual da Paraíba
Rafaela de Albuquerque Paulino
Universidade Estadual da Paraíba
Daiane Medeiros da Silva
Instituto Federal de Educação Ciência e Tecnologia da Paraíba
Alecsandra Ferreira Tomaz
Universidade Estadual da Paraíba
Michelinne Oliveira Machado Dutra
Universidade Estadual da Paraíba
Matheus Pahim Spricido
Universidade Estadual da Paraíba
Maria Eugênia Bresolin Pinto
Universidade Federal de Ciências da Saúde de Porto Alegre
Silvana Santos
silvanasantos@servidor.uepb.edu.br

Universidade Estadual da Paraíba

Research Article

Keywords: professional competence, medical education, formative feedback, family and community medicine, medical residency

Posted Date: October 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11196642/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

Competency-based assessment assumes shared standards for expected performance, yet residents and preceptors may interpret competence, evidence, and progression differently. We examined how evaluative judgments were constructed and negotiated during structured feedback encounters in family and community medicine residency programs. In a qualitative study informed by constructivist grounded theory, 14 resident–preceptor dyads in Paraíba, Brazil, independently rated 34 professional competencies and then justified their judgments in recorded feedback sessions. The 476 paired ratings were descriptively compared, and 456 transcript units were analyzed through initial and focused coding, constant comparison, and analytic memoing. Ratings differed in 56.5% of comparisons, with preceptors generally assigning higher scores than residents assigned themselves. However, score agreement did not necessarily reflect shared reasoning. Residents primarily based judgments on performed activities, practice opportunities, and perceived autonomy, whereas preceptors also drew on theoretical knowledge, overall impressions, expected progression, and anticipated potential, sometimes without direct observation. These findings generated evaluative (mis)alignment as a relational construct describing whether participants' ratings and underlying frames of reference converge or diverge. Feedback became more educationally productive when participants clarified criteria, compared evidence, identified contextual barriers, and translated discussion into actionable plans. The analysis informed S-SKILLS-F, an eight-step framework integrating contextualization, self-assessment, criterion clarification, developmental staging, barrier analysis, perspective integration, planning, and longitudinal follow-up. Evaluative misalignment should therefore be addressed as a negotiated feature of competency assessment rather than as individual inaccuracy. S-SKILLS-F offers an empirically derived approach to aligning assessment, feedback, supervision, and professional development, although its transferability requires evaluation in other contexts.

Introduction

In Brazil, Competency-Based Medical Education (CBME) guides Family and Community Medicine Residency (FCMR) programs. The competency framework defined by the National Commission on Medical Residency (CNRM) comprises 334 items distributed across the two years of training. Although continuous, formative, and contextualized assessment has been established as a cornerstone of residency training (Brasil 2020, 2023), translating this extensive set of competencies into consistent judgments about performance remains challenging. The existence of a competency matrix alone does not ensure that residents and preceptors interpret competencies in the same way or adopt similar criteria for assessing their development. This difficulty stems, in part, from the nature of professional competence itself. Being competent does not merely mean possessing knowledge or performing isolated procedures; it involves mobilizing knowledge, skills, attitudes, values, and professional judgment in response to complex care situations (Costa et al., 2018; Machado et al., 2018; Souza, 2024; Souza et al., 2025; Zafošnik et al., 2024).

Assessing competence therefore requires examining how residents interpret problems, make decisions, and act in real-world primary health care (PHC) settings. However, not all competencies can be directly observed by preceptors, and their descriptions in competency frameworks may be abstract, extensive, or difficult to operationalize. Assessments may therefore draw not only on demonstrated performance but also on general impressions, expectations of progression, or inferences about what residents would be capable of accomplishing (Brito et al., 2025; Campbell et al., 2020; Danilovich et al., 2021; Heim et al., 2023). This tension was highlighted by Souza et al. (2025), who investigated 34 professional competencies in Family and Community Medicine residency programs in Paraíba, Brazil, involving 71 preceptors and 142 residents supervised in clinical settings. Among the residents, the median self-assessment score was 123 points, whereas the median score assigned by preceptors was 140 points ($p < 0.001$). Residents' and preceptors' ratings agreed for 19 of the 34 competencies; for the remaining 15, preceptors assigned higher scores than residents assigned themselves. These findings demonstrated that residents and preceptors produced different evaluations but did not clarify which criteria, experiences, or conceptions of competence underpinned their judgments.

These differences cannot be interpreted solely in terms of the greater or lesser accuracy of residents' self-assessments. Self-awareness, self-efficacy, clinical experience, educational culture, feedback received, and assessment methods may influence both how residents perceive their own performance and how preceptors evaluate it (Eva & Regehr, 2011; Fraundorf et al., 2023; Schultz et al., 2022; Souza et al., 2025). Moreover, different ratings may reflect distinct understandings of the object being assessed, the evidence considered relevant, and the performance expected at each stage of training. When combined with reflection on concrete experiences and evidence-informed external judgments, self-assessment can promote metacognition and self-regulated learning, helping residents recognize their level of autonomy, gaps in their training, and need for supervision (Eva & Regehr, 2011; Fraundorf et al., 2023; Schultz et al., 2022). For self-assessment to serve these purposes, however, the criteria underlying evaluative judgments must be made explicit and discussed.

Although critical and constructive feedback is recognized as an important strategy for competency development in residency training, much of the literature relies on retrospective accounts from residents and preceptors, offering limited access to how judgments are presented, justified, and negotiated during assessment interactions (Burgess et al., 2020; Sargeant et al., 2015; Telio et al., 2015). A gap therefore remains in understanding how and why divergences arise between residents' self-assessments and preceptors' assessments in competency-based assessment processes, as well as how residents and preceptors understand and conceptualize professional competence. The unresolved issue is not simply whether their ratings differ, but what underlies these differences and how they become apparent during assessment interactions. To address this gap, we analyzed structured sessions in which residents and preceptors shared their assessments, explained the rationale for their ratings, and discussed points of agreement and divergence. The findings were also used to develop a structured model of competency-based, learning-oriented feedback. Accordingly, this study was guided by the following research question: How do residents and preceptors conceptualize professional competence, and how do divergences in their evaluative judgments emerge and become negotiated during structured feedback sessions in Family and Community Medicine Residency (FCMR) programs?

Methods

Study design, setting, and participants

We conducted a qualitative study informed by constructivist grounded theory (Charmaz, 2009, 2017; King et al., 2025), an approach suited to examining relational processes and developing explanatory constructs from participants' interactions. Reporting was guided by the Consolidated Criteria for Reporting Qualitative Research (COREQ) (Tong et al., 2007). Data were generated through a structured assessment and feedback procedure in which residents and preceptors independently evaluated professional competencies and then met to compare, justify, and discuss their judgments. The recordings were analyzed to examine how evaluative meanings were constructed and negotiated. The procedure was designed to make evaluative reasoning observable rather than to test the effectiveness of the feedback sessions or of the framework subsequently developed.

In Brazil, Family and Community Medicine Residency (FCMR) is a two-year *lato sensu* graduate program conducted primarily in workplace settings. Programs are provided by higher education institutions, public health schools, and municipal or state administrations in coordination with the Unified Health System (SUS). Their expansion has been associated with policies aimed at strengthening primary health care and extending medical training to rural areas (Barrêto et al., 2019; Brasil, 2015; Separavich & Couto, 2021; Soares et al., 2018). At the time of the study, ten FCMR programs were distributed across the three health macroregions of Paraíba, centered around João Pessoa, Campina Grande, and Patos.

Residents and preceptors affiliated with these programs were recruited through purposive convenience sampling. Participation required agreement from both members of each resident–preceptor dyad. Of the 54 eligible preceptors contacted by telephone or messaging application, 14 agreed to participate with their respective residents. The final sample

comprised 14 dyads: 14 preceptors, 10 first-year residents (R1), and four second-year residents (R2). Five dyads were from Macroregion I, five from Macroregion II, and four from Macroregion III. Because participation was lower than initially anticipated, recruitment required repeated contacts to explain the study, answer questions, and schedule the activities.

Data generation and corpus preparation

Before data generation, each dyad participated in a remote orientation meeting with researchers via Google Meet. During this meeting, they explained the study objectives, the stages of the assessment and feedback procedure, and the use of OBS Studio to record the session. Once participants confirmed that they understood the procedures and could operate the software, they began a new meeting without the researcher present, thereby reducing direct researcher influence on the interaction.

The assessment instrument, adapted from Souza et al. (2025), comprised 34 competencies organized into six dimensions: primary care case management, communication skills, professional commitment, performance of procedures, community-based approach, and practice-based medical knowledge. Each competency was rated on a five-point scale representing the resident's level of confidence, autonomy, and mastery. The development of the instrument is described in Souza et al. (2025).

Residents and preceptors initially rated each competency independently, without access to one another's responses. They then compared their ratings competency by competency. When ratings agreed, participants were instructed to explain the criteria supporting their judgments. When ratings differed, they were asked to discuss the reasons for the discrepancy and propose strategies for developing the competency. Preceptors were also instructed to formulate specific recommendations addressing the identified learning needs.

The recorded sessions lasted between one and three hours. Thirteen dyads discussed all 34 competencies. In one session, a technical problem limited recovery of the recording to discussions of 14 competencies. The recordings were uploaded to a restricted-access folder and initially transcribed using TurboScribe. The researchers subsequently reviewed each transcript while listening to the complete recording, a process that took an average of 2.5 hours per session. Unclear passages were checked up to five times, and passages that remained unintelligible were marked with a question mark. Names of individuals, institutions, health services, and locations were replaced with generic descriptors. Dyads, preceptors, residents, and competencies were assigned alphanumeric codes, such as D01, P01, R01, and C01. Dyad numbers were randomly assigned and did not correspond to the order in which the sessions were recorded.

Data analysis and reflexivity

The analysis was conducted without specialized qualitative data-analysis software. Two researchers initially examined the material independently using initial coding, focused coding, constant comparison, and analytic memoing. One researcher (SS) had previous experience in qualitative research and contributed to data interpretation. Divergent codes and interpretations were subsequently reviewed jointly and discussed until consensus was reached. Consistent with the constructivist orientation of the study, the analytical categories were understood as being developed through the researchers' engagement with the recordings, transcripts, comparisons, and analytic memos rather than as direct representations extracted from the data.

The unit of analysis was the discussion of an individual competency and included the resident's self-assessment, the preceptor's assessment, and the rationales provided by both participants. The corpus comprised 456 units: 442 from the 13 complete recordings and 14 from the partially recovered recording. Counts of rating agreement and disagreement were used descriptively to characterize the corpus and guide qualitative comparisons; they were not treated as inferential evidence of assessment accuracy.

During initial coding, the transcripts were examined line by line to identify actions, assessment criteria, conceptions of competence, interpretations of the rating scale, and evidence used to support evaluative judgments. Each unit was compared with other units from the same dyad and subsequently with units from other dyads. This process made it possible to identify recurrences, variations, and tensions in how residents and preceptors evaluated competencies. At the same time, analytic memos recorded interpretive hypotheses and provisional relationships among codes, thereby guiding subsequent comparisons and the refinement of the analytical categories.

Through this process, the construct of evaluative (mis)alignment was developed to describe the relationship between the ratings assigned by residents and preceptors and the frames of reference used to justify them. Alignment was not determined solely by numerical agreement. Identical ratings could be supported by different assessment criteria, whereas different ratings could arise from distinct conceptions of the competency, the evidence considered relevant, or the performance expected at a particular stage of training.

To examine these relationships systematically, the units were initially classified into four forms of agreement. Strong agreement (C+) occurred when the ratings were identical and supported by convergent justifications and observable evidence. Divergent agreement (C−) occurred when the ratings were identical but residents and preceptors used different criteria to justify them. Agreement by inference (CI) referred to identical ratings assigned without direct observation and based instead on assumptions about the resident's ability. Agreement without justification (C0) was recorded when identical ratings were not accompanied by sufficiently clear explanations.

Differences between ratings were also classified into four forms. Positive disagreement (D+) occurred when the preceptor assigned a higher rating than the resident's self-assessment. Negative disagreement (D−) occurred when the preceptor assigned a lower rating than the resident. Inference-based disagreement (DI) was identified when the difference between ratings was supported by assumptions about performance rather than observable evidence. Unjustified disagreement (D0) was recorded when participants did not provide sufficiently clear reasons for the different ratings.

During focused coding, the justifications supporting the ratings were compared to identify the frames of reference used to construct the evaluations. The codes were progressively grouped into four analytical axes: conceptions of professional competence; conceptions of assessment and uses of the rating scale; conditions of observation and opportunities for learning; and approaches to feedback.

The first axis examined how participants defined professional competence. Competence could be associated with theoretical knowledge, practical performance, the ability to explain or teach, or the resident's presumed potential. The second axis addressed conceptions of assessment and interpretations of the rating scale. Ratings could be guided by demonstrated progression, expectations regarding future performance, motivational purposes, or symbolic and moral meanings attributed to particular scores. The third axis examined the conditions under which judgments were produced, including direct observation, lack of opportunities to perform a particular activity, and inferences about what the resident would be capable of accomplishing. The fourth axis examined how feedback was provided, distinguishing more assertive approaches - characterized by explicit criteria, problem-focused discussion, and actionable guidance - from less assertive approaches that were predominantly validating, generic, or inferential. The analytical process can be illustrated by unit D03C01, which concerned the Person-Centered Clinical Method (PCCM):

R03: "I think I assess correctly; I use the person-centered clinical method."

P03: "I rated you a 4 because I think you do very well; you really have this competency. But I think there are some aspects of the person-centered clinical method that [...] theorizing about this might be something that still needs to be explored in greater depth [...] given that you are in this transition from R1 to R2."

The unit was initially classified as negative disagreement (D−) because the resident assigned a higher self-rating than the rating assigned by the preceptor. During focused coding, the resident's justification was associated with the practical application of the competency, expressed in the statement "I use the person-centered clinical method." The preceptor, however, considered theoretical mastery, the ability to explain or teach the method, and the performance expected during the transition from the first to the second year of residency. The analytic memo prepared for this unit noted that the difference arose not only from the numerical ratings but also from the frames of reference used to define competence: the resident emphasized practical application, whereas the preceptor emphasized theoretical mastery, the ability to teach, and expectations regarding educational progression. Comparison of this case with other units revealed the recurrence of these forms of judgment and supported further refinement of the analytical categories.

In the final stage of analysis, the categories developed through focused coding and constant comparison were integrated to identify the conditions under which feedback became more educationally productive. Particular attention was given to interactions in which participants contextualized the assessment, compared self-assessment with external judgments, clarified criteria, identified the resident's developmental stage, examined barriers and learning opportunities, integrated different perspectives, formulated actionable plans, and considered the need for longitudinal follow-up. Organizing these emergent analytical elements into a sequence of educational actions informed the development of the provisional S-SKILLS-F framework for competency-based, learning-oriented feedback. The framework was therefore derived from the observed assessment and feedback processes; its effectiveness was not evaluated in the present study.

Results

The analysis yielded the central construct of evaluative (mis)alignment, understood as a relational and multifactorial phenomenon. In addition to the agreement or disagreement between scores, the construct incorporated conceptions of competence, judgment criteria, evidence of practice, interpretations of the scale, modes of providing feedback, and specific training conditions.

A total of 476 paired rating comparisons were obtained between residents and preceptors. Ratings agreed in 207 comparisons (43.5%) and differed in 269 (56.5%). Of these comparisons, 456 included narrative content suitable for analysis of the underlying rationales: 442 from 13 complete recordings and 14 recovered from a recording affected by a technical failure. The remaining 20 comparisons retained the ratings recorded on the instrument but lacked sufficient narrative content for analysis. Category D+, in which the preceptor assigned a higher rating than the resident, predominated. However, numerical agreement did not necessarily indicate evaluative alignment: some identical ratings were supported by different criteria or by inference rather than shared evidence, whereas different ratings sometimes reflected different definitions of the competency under discussion (Fig. 1).

The ongoing comparison among the units showed that these patterns were produced by four interrelated processes: different conceptions of what it means to be competent; different ways of interpreting and applying the assessment scale; distinct relational approaches to providing feedback; and tensions between the competencies outlined in the curriculum and the actual learning opportunities in the field. While highlighting these differences, the meetings facilitated reflection, the negotiation of meanings, and learning planning. The integration of these processes gave rise to the S-SKILLS-F model, presented at the end of the results.

(Mis)alignment between Doing, Knowing, and Potential

Among the residents, a prevailing understanding of competence was associated with doing, concrete performance, and the confidence built through experience. Their self-assessments were based primarily on what they accomplished on the job, the repetition of activities, and the autonomy they developed in their daily care work. This conception is evident in statements such as "I know that I apply [my knowledge] appropriately and that I am also able to teach others" (D05C01), "I

consider myself someone who requests additional tests rationally" (D05C04), and "I feel completely confident in this area" (D12C05). In these cases, the assigned rating reflected an assessment of a practice that had actually been carried out.

The lack of concrete experience also influenced these judgments. When they had not performed a particular activity, the residents recognized the need for supervision and assigned lower ratings. When referring to the intrauterine device (IUD), one resident stated: "I gave it a 1 because I was unable to insert an IUD" (D03C23). Others reported, "I've never done it" (D02C26) and "I gave them all a score of 1 because it's something I haven't done yet during my residency" (D09C24-28). Expressions such as "I've never done it," "I haven't had any practice," and "I can't perform it" described not only subjective insecurity but also a lack of opportunities to apply their skills in real-world situations.

Among the preceptors, assessments often incorporated other elements. Even without direct observation, they considered theoretical and procedural knowledge, their overall impression of the resident, and the expectation that the resident would be able to perform the activity when an opportunity arose. This shift from observed practice to future potential is evident when the preceptor states: "I rated you as 'capable' because I believe we are capable of performing the procedure as long as we learn about it" (D01C21). In another situation, competence was presumed based on the resident's knowledge: "I consider you capable [...] I just didn't mark 'confident' because I haven't seen you perform any" (D01C07).

The assessment was also based on analogy with other performances: "I didn't observe the other procedures, but I'm sure that given your ease in performing the otoscopy, you're capable of doing the other things" (D06C02). In "I gave you a five; you're on call... there's always an abscess to drain" (D07C27), the assessment was based on an expectation regarding experiences likely encountered outside of direct supervision, rather than on performance observed by the preceptor.

The difference between the assessment criteria became particularly evident in pair 9. The resident justified the lower scores by citing a lack of practice: "I gave them all a 1 because it's something I haven't done yet during my residency." The preceptor, however, assigned a higher score based on presumed ability: "I gave a 4; I think they're capable - now they just need to gain the practical experience" (D09C24-28). The same task was thus evaluated according to different logics: the resident judged what they were able to perform; the preceptor, what they believed the resident would be able to accomplish.

This tension between doing, knowing, and potential contributed to the predominance of category D+. The higher scores assigned by the preceptors did not necessarily reflect greater knowledge of the residents' performance. In some units, they reflected a broadening of the scope of what was being assessed, which shifted from the demonstrated competency to include knowledge, expectations for progress, and future potential.

The interpretation of the scale further deepened these differences. Although conceived as a progressive description of safety, autonomy, and mastery, the scale was used in accordance with this purpose in 41.8% of the preceptors' evaluations and in 52.9% of the residents' self-evaluations. In the remaining cases, the scores were reinterpreted according to numerical, classificatory, moral, or motivational logic.

Rating 1, which was intended to indicate a greater need for supervision and pedagogical support, was associated with failure, punishment, or harm: "a score of one would cause harm" (D05C09). In another interaction, a low score was perceived as a threat to established trust: "Why? Such a low score? I trusted you so much" (D13C31). In these situations, the scale ceased to indicate a learning need and came to be interpreted as a personal judgment of the resident.

The ratings were also used for motivational purposes. In the case of the *Projeto Terapêutico Singular* (PTS; Singular Therapeutic Project), a collaborative, individualized care-planning approach used in Brazil, even without observing performance, preceptors assigned intermediate or high levels as a form of encouragement: "I rated them as 'capable' [...] the STP is a bit difficult to evaluate because we haven't worked on it yet" (D04C07). In another interaction, the following was heard: "I gave everyone a 3 because I didn't see any" (D13C25). The concept expressed less the demonstrated stage of competence and more the desire to encourage the resident or recognize their potential.

At the upper end, level 5 was interpreted as perfection or excessive pretension, and not necessarily as autonomy associated with the ability to teach: “I think that would be a bit arrogant” (D13C01), “I don’t think it makes sense for me to feel completely confident teaching that” (D01C13) and “I don’t think even I would reach a 5 there” (D10C12). The symbolic weight attributed to the extremes limited the scale’s educational use and concentrated part of the evaluations on the intermediate ratings.

Numerical agreement could also mask conceptual differences. In one unit, the resident justified the score based on experience: “I performed suturing, canthoplasty, and abscess drainage,” while the preceptor assigned the same score despite not having observed the procedures: “I gave a 3 to all of them because I didn’t see any of them” (D13C28). Therefore, identical scores did not guarantee that the resident and preceptor shared the same definition of competency or considered the same evidence.

Ways of Conducting Feedback

Differences in judgment did not arise only before the sessions. They were reinforced, made explicit, or reworked depending on how the feedback was conducted. Preceptorship approaches were understood as relational ways of organizing listening, presenting evaluations, addressing performance issues, acknowledging difficulties, and developing learning strategies.

In interactions classified as assertive preceptorship, preceptors made their criteria explicit, related concepts to concrete situations, and formulated actionable guidance. The feedback combined recognition of potential, identification of weaknesses, and planning for opportunities for development. Guidance such as “gather the team” and “plan the waiting room” transformed general difficulties into actions that could be carried out in the clinical setting.

These encounters also allowed for greater openness, enabling the preceptor to revise their interpretation in light of the resident’s explanation. Interactional markers such as “ummm” indicated moments of pause, elaboration, and reorganization of understanding. Rather than merely communicating a judgment, the resident and preceptor compared evidence, clarified the meaning of competencies, and arrived at a shared interpretation of performance.

In interactions classified as non-assertive preceptorship, validation, inferences, and projections about the future predominated. Expressions such as “you’ll learn,” “you’ll get the hang of it over time,” and “we’ll work on this in R2” were reassuring to the resident but shifted the discussion from the current competency to an expectation of future development. This shift aligned with the D + and CI categories, as the judgment was based more on confidence in future ability than on observation of performance.

In other situations, difficulties were acknowledged through statements such as “this really is difficult” and “it’s a challenge for us,” without exploring their causes or formulating strategies to overcome them. Validation helped preserve the pedagogical bond, but, on its own, it did not transform the acknowledged difficulty into a concrete learning opportunity.

Across the analyzed interactions, less assertive approaches were associated with inferential judgments, higher competency ratings, and agreements based on assumptions. More assertive approaches made the evidence and criteria available for discussion and more often produced actionable guidance. These patterns should be interpreted as qualitative relationships within this dataset rather than estimates of the comparative effectiveness of different feedback approaches.

The relational conditions and the order in which judgments were presented also shaped the interaction. In the analyzed sessions, when the preceptor presented a judgment first, the resident's subsequent account was more likely to adopt the evaluator's frame of reference. When the resident began with self-assessment, their experiences, uncertainties, and criteria became available for comparison before the preceptor's judgment was introduced. This pattern supports treating conversational order as a mechanism to be examined rather than as proof of a causal effect.

These findings showed that the instrument, by itself, did not guarantee formative feedback. Its potential depended on creating a space in which the resident's perspective could be presented before being influenced by external judgment, criteria could be made explicit, differences could be explored, and conclusions could be transformed into development strategies.

Between Curriculum and Practice

As residents and preceptors explained the concepts, the meetings revealed competencies outlined in the curriculum but rarely incorporated into practice. This (mis)alignment did not stem solely from differing perspectives. In several units, it reflected the gap between what the program expected residents to develop and the opportunities actually available in the clinical settings.

The disconnect between knowledge and practice was particularly evident in the STP - individualized care plan: "I've never carried out an STP" (D01C07), "it's a tool that you haven't been able to apply in practice" (D03C07), "the STP we haven't worked on yet" (D04C07), and "we've never really stopped to carry out that part of the STP" (D12C07). Residents and preceptors recognized its importance but described few experiences of jointly developing and providing longitudinal follow-up for a STP.

A similar situation occurred with family-centered approach tools: "I've never actually drawn a genogram or an eco-map" (D01C10), "I don't do genograms or eco-maps at all" (D03C10), and "I know I should include them more, but they don't really come into practice that much" (D11C10). Conceptual knowledge did not regularly translate into supervised practice.

Regarding procedural competencies, the lack of opportunities, materials, and supervision limited the transition from technical knowledge to implementation: "I've never done it" (D02C26), "there are no materials" (D06C24), "I know the technique, but I've never actually done it" (D13C21), and "since I do it so rarely, I gave it a 3" (D08C22). In these cases, the lower scores could not be interpreted solely as individual deficiencies, as they also reflected the conditions available for learning and practicing.

Home visits revealed difficulties with planning and low resident participation in organizing the process: "I try, but I can't do it" (D03C09), "this issue of planning and organizing the team is something I struggle with a lot" (D05C09), and "the health workers are the ones who plan the itinerary" (D07C09). Conducting the visits did not necessarily mean developing the ability to plan, prioritize, and coordinate home care.

Territorialization also remained largely unimplemented: "I've never done any very visual, graphic, or digital activities related to territorialization" (D01C29), "I've only done territorialization on paper, but I've put zero of it into practice" (D05C29), and "it's something I'd know how to do, but I rarely apply it to my practice" (D08C29). This competency was recognized as a component of PHC but was rarely utilized to organize daily care.

In health education, residents reported difficulties with group planning, the use of indicators, and the organization of collective actions: "It's something that's often talked about, but it's not put into practice" (D02C32), "I'm still trying to understand how it's done" (D04C32), and "We still have to structure these groups" (D06C32). Similar weaknesses emerged regarding shared leadership, the implementation of workflows, and the quality of the team's work: "It's not something I do" (D04C34), "I don't feel like I try to encourage it" (D05C34), and "I've never implemented protocols or workflows" (D09C34).

In care coordination, tools such as STP, telehealth, referral, and counter-referral were known but rarely used: "I'm familiar with the tools, but I'm not in the habit of using them" (D01C07), "telehealth is something I haven't used" (D08C07), and "I get little feedback from counter-referrals" (D09C07). Recurring expressions - "there's no time," "no cases come up," "we don't do that at the facility," and "there aren't any materials"- showed that the development of competencies was constrained by work organization, available resources, and opportunities for supervision.

The instrument itself also revealed limitations. Extensive competencies or those composed of different actions made it difficult to assign a single rating: “there were too many very different things grouped together under the same topic” (D04C27), “when you put several things that are so different together, it becomes very difficult to score” (D04C31), and “it brings several elements together” (D05C07). The difficulty in scoring reflected the tension between the complexity of professional performance and its representation by a single category.

The structured meeting thus served as a lens through which to examine the training program. By asking participants to justify their ratings, it brought to light not only individual gaps but also curricular weaknesses, service limitations, and characteristics of the instrument that hindered the development, observation, and assessment of competencies.

Clarification, Metacognition, and Learning

For many participants, the session represented the first structured opportunity to read about and discuss the competencies expected during the residency. The experience was described as “very comprehensive,” capable of “reinforcing the competencies we must develop during the residency,” “thinking through every detail of our practice,” and “discussing and planning improvements to our learning.”

The shared reading also brought doubts and ambiguities to light: “I was unsure about this part” (D03C15), “I didn’t understand whether it referred to group settings, the waiting room, or during individual consultations” (D05C15), and “when there are several different things listed, it becomes very difficult to score them” (D04C31). When attempting to assign a score and justify it, participants had to interpret the meaning of the competency and define which practices were being assessed.

Self-assessment helped clarify the experiences that guided the resident’s judgment. It was not limited to describing what the participant knew, but also included what they were able to do, what they were unable to do, or what they had not yet had the opportunity to try. When these justifications were compared to the external evaluation, previously implicit differences could be recognized and reworked.

This dynamic is evident in statements such as “the fact that you’re telling me this really broadens my perspective” (D13C34) and “it was good… to get some sense of it” (D05C34). The comparison of perspectives fostered awareness of one’s own performance, gaps in knowledge, and the criteria used in the evaluation. Feedback thus shifted from the mere communication of a result to a process of constructing meaning about learning.

However, simple agreement did not guarantee reflection. In brief interactions, “I consider myself completely confident” was answered only with “I also marked ‘completely confident’” (D01C01); in another, “I scored it as a 5” was followed by “we agreed, didn’t we?” (D02C01). Although the ratings were the same, the evidence, the definition of competence, and the possibilities for progression remained largely implicit.

The educational potential of the session therefore depended on the ability to go beyond a mere comparison of concepts. When residents and preceptors clarified what they understood by “competence,” what evidence they used, why a gap persisted, and what could be done, the feedback served simultaneously as an assessment tool, a metacognitive space, and a learning planning instrument.

From Analysis to the S-SKILLS-F Model

Across the four analytic dimensions, difficulties in competency assessment appeared to be connected rather than isolated. Residents and preceptors sometimes entered evaluation meetings with different understandings of the instrument, the meaning of a competency, or the evidence required to justify a rating. The resident’s account could also be shaped by hearing the preceptor’s judgment first. Once differences in assessment emerged, the discussion did not always establish whether they reflected different interpretations of the criteria, different evidence of performance, or differences between

what the resident could explain and what they had demonstrated in practice. Learning opportunities and constraints within the clinical setting were not consistently considered, and identified needs did not always lead to actions that could be reviewed later.

The S-SKILLS-F model was developed from these findings to structure feedback meetings for the assessment of professional competencies in family and community medicine residency. It brings together the resident's self-assessment and the preceptor's assessment in a conversation centred on the evidence and reasoning behind each judgment. Its purpose is not simply to produce an agreed rating. By making alignments and misalignments visible, the meeting can establish a more precise understanding of current performance, the support needed for further development, and the next opportunities to learn. The model is presented in Fig. 2.

The process begins before the meeting. Resident and preceptor need to know which competency or entrustable professional activity (EPA) will be discussed, which assessment instrument will be used, and what its criteria and performance levels mean. This preparation gives the resident a basis for self-assessment and helps both participants enter the meeting with a common understanding of its purpose. During the encounter, the resident presents their assessment first, including examples from practice and areas of uncertainty. The preceptor then shares their assessment and the observations on which it rests. This order preserves the resident's initial perspective and allows both accounts to be examined, rather than treating the preceptor's rating as the starting point for the conversation.

The central work of the meeting is to interpret the two assessments together. A difference in ratings may point to a gap in performance, but it may also reveal that resident and preceptor have used different criteria, relied on different clinical experiences, or understood the expected degree of autonomy differently. They therefore clarify what the competency requires in practice, what the available evidence shows, and where further observation is needed. They also consider factors that may have limited the resident's opportunity to demonstrate or develop the competency, including access to relevant cases, supervision, time, resources, and relational or emotional challenges. This discussion helps locate the resident's current developmental needs without attributing every difficulty to the individual.

The meeting should end with a shared account of the priorities for development and a plan that can be acted on in the clinical setting. Resident and preceptor identify what the resident will practise, what support or opportunities the preceptor or programme will provide, what evidence of progress they will seek, and when they will review it. Follow-up is part of the model: subsequent encounters allow them to examine new evidence, assess whether the planned actions were feasible, and revise their understanding or goals as the resident develops. The stages thus guide a continuing cycle of assessment and learning, rather than a single feedback conversation.

To support judgments about progression, S-SKILLS-F also proposes five developmental levels (Fig. 3). These are applied to the specific competency or EPA under discussion, rather than used as an overall label for the resident. They prompt consideration of how the resident mobilises the competency in clinical practice, how consistently this occurs, what supervision is needed, and how the resident reflects on and adjusts their performance. Used in this way, a level is a provisional description grounded in available evidence. It helps resident and preceptor discuss appropriate supervision and a realistic next step, while leaving room for reassessment as further opportunities to perform arise.

S-SKILLS-F was formulated after the study's data analysis. It did not guide data collection, and its implementation and effects were not evaluated in this study. It is therefore presented as a proposed pedagogical model derived from the identified challenges in competency assessment and feedback.

Discussion

The main innovation of this study was to use recordings of structured competency-assessment and feedback sessions as empirical data and to analyze how judgments were presented, justified, negotiated, and, at times, reconstructed. A recent

review showed that much research still describes experiences and perceptions, while fewer studies examine how evaluative information is converted into action (Fuentes-Cimma et al., 2024). Previous observational studies have focused primarily on educators' behaviors and their alignment with best-practice recommendations (Johnson et al., 2019). By combining independent evaluations with recorded interactions, this study provided access to the reasoning expressed during the assessment encounter, rather than relying only on retrospective accounts of feedback.

This methodological approach revealed that (mis)alignment cannot be reduced to the accuracy of self-assessment or the preceptor's ability to assign the "correct" grade. The literature already recognizes that judgments in clinical situations are interpretive and influenced by the evaluator, the trainee, the established relationship, and the context of practice (Gingerich et al., 2014; Govaerts & Van der Vleuten, 2013; Lingard, 2009; Watling & Ginsburg, 2019). The central conceptual contribution of this study is evaluative (mis)alignment: the relationship between a rating and the reference points used to justify it. Divergence can begin before performance is assessed: residents and preceptors do not always evaluate the same object, even though they use the same competency and scale. While one judgment is grounded in practice actually performed, the other may incorporate theoretical knowledge, an overall impression, and presumed potential. Conversely, identical ratings need not indicate a shared understanding of performance.

This finding challenges an implicit premise in many CBME instruments: that standardized descriptions would be sufficient to produce a shared understanding of expected performance. Competence, however, remains a polysemic concept, capable of being associated with knowledge, task execution, autonomy, the ability to teach, or the contextual mobilization of different professional resources (Alharbi, 2024; Carraccio et al., 2017; Englander et al., 2017; Epstein & Hundert, 2002; Frank et al., 2010; Ten Cate & Scheele, 2007). This polysemy is not merely a conceptual debate: it influences which evidence participants regard as relevant. A numerically consistent evaluation may conceal incompatible frames of reference, while a disagreement may reflect different understandings of what counts as competence rather than different appraisals of the same performance.

The rating scales were not neutral tools. Their levels took on moral, relational, and motivational meanings: lower ratings could represent failure or a threat to the relationship, while higher ones were associated with perfection or personal recognition. Variability, therefore, stemmed not only from the observation or processing of evidence but also from the anticipation of the emotional and pedagogical effects of each rating. A rating chosen partly to reassure or encourage a resident may sustain the conversation, yet obscure the supervision and learning needs that assessment is intended to identify.

This mechanism helps explain why the implementation of competency-based instruments does not always result in assessment for learning. In a recent study of residents in Canadian programs, Hempel et al. (2026) observed that assessments of professional activities were perceived as bureaucratic exercises when they did not yield meaningful feedback. The authors caution that emphasizing individual attitudes exclusively may place the blame on the resident for problems generated by the evaluation system itself. The present study adds that even a completed assessment can have limited educational value when ratings are assigned without observation, criteria remain implicit, or presumed potential substitutes for evidence of practice.

The way the session was conducted could either exacerbate or mitigate this problem. Contemporary literature has shifted the concept of feedback from information conveyed by the instructor to a process of meaning-making, in which the learner interprets, discusses, and uses information to modify performance. Despite this shift, residents still report feedback that is unclear or rarely accompanied by actionable plans (Shafian et al., 2024). Observational research has also found limited learner involvement in self-assessment, action planning, and follow-up (Johnson et al., 2019). The recordings in this study showed how early validation, the preceptor's presentation of a judgment before the resident's account, and assumptions about the resident's potential could close down inquiry into the evidence and criteria behind an assessment.

The order of the conversation proved particularly relevant. When the preceptor presented their evaluation first, the asymmetry of the relationship could encourage the resident to align their account with that judgment. Beginning with self-assessment does not mean assigning the resident responsibility for producing a precise measure of performance. It preserves their perspective as material for discussion alongside the preceptor's observations, including the experiences, doubts, and criteria informing each account. This interpretation aligns with the reformulation of self-assessment proposed by Eva and Regehr (2011), who understand it as self-monitoring and a guided search for external information. A study of co-created narrative feedback similarly associated learner participation with reflection, goal-setting, and self-regulation (Cavallaro et al., 2026). Our contribution is to show, in recorded competency-assessment encounters, that self-assessment also exposes the reference points through which residents interpret their practice.

At the same time, performance cannot be separated from opportunities to learn and demonstrate a competency. Recent studies on supervision and trust-building reinforce that supervisor availability, time, psychological safety, explicit reasoning, and organizational conditions influence autonomy and professional development (Lerch et al., 2025). In PHC, limited demonstration of a primary care competency may reflect insufficient access to relevant cases, supervision, resources, or opportunities for practice, as well as an individual learning need. A low rating should therefore prompt inquiry into both performance and its conditions. Assessment may expose a gap between the intended curriculum and what the workplace has enabled the resident to practise or show.

Studies have described difficulties in implementing family-centered care, interprofessional collaboration, person-centered practice, and care coordination in primary health care (Campos & Domitti, 2007; Ceccim & Feuerwerker, 2004; Gantayet-Mathur et al., 2022; Matuda et al., 2015; Švab & Cerovečki, 2024). Our findings connect these educational conditions to an assessment problem: where direct evidence of practice is scarce, a preceptor may fill the gap with an inference about what the resident would know or be able to do. Such an inference can conceal, rather than identify, a need for teaching, practice, or further observation.

The complexity of the competency matrix intensifies this tension. Extensive matrices seek to represent the breadth of professional practice, but they can fragment the work and burden assessment (Schuwirth & Van der Vleuten, 2011; Ten Cate, 2017; Van der Vleuten et al., 2012). In this study, even the reorganization of the national matrix into 34 summary competencies required lengthy meetings and created difficulties when distinct actions were grouped under a single item. Changing the number of competencies alone cannot resolve differences in their interpretation; residents and preceptors also need space to discuss their meaning, supporting evidence, and progression.

It was precisely from the interplay between these problems that S-SKILLS-F emerged. The model was not designed in advance to organize the meetings nor applied as an intervention during data collection. It was constructed afterward, through the integration of the categories identified in the analysis. It supports structured meetings in which resident and preceptor bring their assessments into dialogue, clarify the criteria and evidence behind them, examine developmental and contextual needs, agree on actions, and revisit progress. Its distinctive emphasis is on negotiating the foundations of evaluative judgment before using that judgment to plan development.

S-SKILLS-F shares with the Relationship, Reaction, Content, and Coaching (R2C2) model an understanding of feedback as a dialogic and relational process (Sargeant et al., 2015; Sargeant et al., 2018; Thakur et al., 2023). It addresses a particular challenge in competency assessment: participants may need to clarify what they mean by the competency, which observations support their judgments, and whether a rating describes demonstrated performance or presumed potential before they can identify a useful next step.

The sequential organization of S-SKILLS-F was inspired by SPIKES (Baile et al., 2000), which structures complex conversations in clinical practice. The similarity lies in organizing a potentially difficult interaction into recognizable and interdependent stages, not in transposing its content. The stages of S-SKILLS-F were derived from the present analysis; the model's feasibility and educational effects remain to be tested.

The implications go beyond adopting a new script. If judgments are relational and contextual, high-quality assessment requires pedagogical training for preceptors, opportunities for direct observation, dedicated time, shared criteria, and an institutional culture that allows gaps to be recognized without equating them with personal failure. Faculty development can improve feedback practices, but its effects require institutional reinforcement (Mukhia & Tenzin, 2025). S-SKILLS-F may help participants make these conditions visible during an encounter; it cannot itself provide the observation, supervision, or learning opportunities on which credible assessment depends. Although developed in Brazilian FCM residency programs, its underlying questions about implicit criteria, limited evidence, relational pressures, and unequal learning opportunities warrant study in other workplace-based assessment settings.

This study has limitations. Its grounded theory-informed analysis used selected procedures, including coding, constant comparison, and analytic memoing; the categories should be understood as interpretations developed in this research context rather than a universal account. The study was conducted in FCM residency programs in one Brazilian state and involved predominantly first-year residents. Participation depended on both members of a resident–preceptor pair, which may have introduced selection bias. The sessions were structured for research and may have differed from routine feedback encounters; the recordings captured participants' assessments and discussion; not independent direct observation of the clinical performances being rated. A technical problem prevented retrieval of the justifications for 20 comparisons, although the ratings remained available. Neither S-SKILLS-F nor its proposed developmental levels were implemented or validated. Future research should examine feasibility, acceptability, use in practice, and consequences for feedback, supervision decisions, and competency development.

Conclusion

This study demonstrated that evaluative (mis)alignment between residents and preceptors extends beyond differences in scores. By analyzing recorded assessment encounters alongside independent ratings, it identified how definitions of competence, accepted evidence, relational considerations, and opportunities for practice shaped the judgments expressed during feedback. Numerical agreement did not necessarily indicate shared reasoning, and disagreement did not necessarily indicate inaccurate judgment by either participant.

Based on these findings, S-SKILLS-F structures assessment as a dialogic and longitudinal process. Its contribution is to make comparison and negotiation of evaluative reference points part of the encounter, connecting that work to learning plans and follow-up. The model offers a pedagogical response to the challenges identified in FCM residency, but its implementation and educational effects remain to be evaluated.

Abbreviations

CAAE
Certificate of Ethical Review
CBME
Competency—Based Medical Education
CNRM
National Commission on Medical Residency
COREQ
Consolidated Criteria for Reporting Qualitative Research
FCM
Family and Community Medicine
FCMR
Family and Community Medicine Residency
GT

Constructivist Grounded Theory
IUD
Intrauterine Device
PCCM
Person—Centered Clinical Method
PHC
Primary Health Care
R1
First—year resident
R2
Second—year resident
R2C2
Relationship, Reaction, Content, and Coaching
S
SKILLS—F—Setting, Self—assessment, Key, Identification, Learning Barriers, Linking Perspectives, Strategy, and Follow—up Model
SPIKES
Setting, Perception, Invitation, Knowledge, Emotions, and Strategy/Summary
STP
Singular Therapeutic Project
SUS
Unified Health System

Declarations

Ethical Approval and Consent to Participate

This study was conducted as part of the project "Educational Intervention: Competency-Based Training in the Family and Community Medicine Residency and the Multidisciplinary Family Health Residency in the State of Paraíba," approved by the Research Ethics Committee of the State University of Paraíba (CAAE No. 73484623.5.0000.5187). The study complied with Brazilian ethical standards for research involving human participants, including National Health Council Resolutions No. 466/2012 and No. 510/2016 and Operational Standard No. 001/2013. Written informed consent to participate was obtained from all residents and preceptors before data collection.

Consent for Publication

The manuscript presents only anonymized data and quotations, without any information that would allow for the direct identification of participants. All participants provided written consent for the scientific use and anonymized publication of the data and statements produced during the study.

Competing interests

The authors declare that they have no financial or non-financial competing interests related to this work.

Use of Generative Artificial Intelligence

ChatGPT (OpenAI) was used to assist with the visual creation of the figures.

Authors details

1 Universidade Estadual da Paraíba (UEPB), João Pessoa, Brazil

2 Fundo Municipal de Saúde de Campina Grande (FMS), Campina Grande, Brazil

3 Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB), João Pessoa, Brazil

4 Escola de Saúde Pública da Paraíba (ESP-PB), João Pessoa, Paraíba, Brazil

5 Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, Brazil

Funding

The research received no specific grant from any funding agency. The article processing charge (APC) for open access publication was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) through the CAPES–Springer Nature agreement.

Author Contribution

SS, HOQB, and AFT conceived the study, collected the data, and conducted the analysis. RAP, WPS, and ENS contributed to participant recruitment and reviewed the findings. DMS and MEBP critically reviewed the manuscript. MOMD reviewed the final version of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank the residents and preceptors who participated in the study and shared their experiences and reflections during the evaluation and feedback sessions.

Availability of Data and Materials

The recordings and full transcripts are not publicly available because they contain contextual information that could allow for the indirect identification of participants and residency programs. Duly anonymized analytical materials may be made available by the corresponding author upon reasonable request, subject to the preservation of confidentiality and the terms of informed consent.

References

1. Alharbi, N. S. (2024). Evaluating competency-based medical education: a systematized review of current practices. *BMC Medical Education*, *24*, 612. https://doi.org/10.1186/s12909-024-05609-6
2. Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES-A six-step protocol for delivering bad news: application to the patient with cancer. *The Oncologist*, *5*(4), 302–311. https://doi.org/10.1634/theoncologist.5-4-302
3. Barrêto, D. S., Neto, M., Figueiredo, A. J., Sampaio, A. M., Gomes, J., L. B., & Soares, R. S. (2019). The More Doctors Program and Family and Community Medicine residencies: articulated strategies of expansion and interiorization of medical education. *Interface - Comunicação, Saúde, Educação*, *23*(Suppl 1), e180032. https://doi.org/10.1590/Interface.180032

4. Brasil. Ministério da Educação. Comissão Nacional de Residência Médica (2020). *Resolução CNRM nº 9, de 30 de dezembro de 2020: Aprova a Matriz de Competências dos Programas de Residência Médica em Medicina de Família e Comunidade*. http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=119641-11-matriz-de-competencias-em-medicina-familia-e-comunidade&category_slug=agosto-2019-pdf&Itemid=3019
5. Brasil (2023). Ministério da Educação. Comissão Nacional de Residência Médica. *Resolução nº 4, de 1º de novembro de 2023: Dispõe sobre os procedimentos de avaliação dos médicos residentes e dá outras providências*. https://www.gov.br/mec/pt-br/residencia-medica/pdf/RESOLUON4DE1DENOVEMBRODE2023RESOLUON4DE1DENOVEMBRODE2023DOUImprensaNacional.pdf
6. Brasil. Ministério da Saúde (2015). *Programa Mais Médicos: dois anos - mais saúde para os brasileiros*. https://bvsms.saude.gov.br/bvs/publicacoes/programa_mais_medicos_dois_anos.pdf
7. Brito, L. A. R., Caramori, U., D'Abreu, N. B., Filho, C., Cunha, C., L. O., & Bragança, J. F. (2025). Atividades profissionais confiáveis para a residência em Medicina de Família e Comunidade no contexto brasileiro. *Revista Brasileira de Educação Médica*, *49*(1), e029. https://doi.org/10.1590/1981-5271v49.1-2024-0198
8. Burgess, A., van Diggele, C., Roberts, C., & Mellis, C. (2020). Feedback in the clinical setting. *BMC Medical Education*, *20*(Suppl 2), 460. https://doi.org/10.1186/s12909-020-02280-5
9. Campbell, C., Hendry, P., Delva, D., Danilovich, N., & Kitto, S. (2020). Implementing competency-based medical education in family medicine: a scoping review on residency programs and family practices in Canada and the United States. *Family Medicine*, *52*(4), 246–254. https://doi.org/10.22454/FamMed.2020.594402
10. Campos, G. W. S., & Domitti, A. C. (2007). Apoio matricial e equipe de referência: uma metodologia para gestão do trabalho interdisciplinar em saúde. *Cadernos de Saúde Pública*, *23*(2), 399–407. https://doi.org/10.1590/S0102-311X2007000200016
11. Carraccio, C., Englander, R., Gilhooly, J., Mink, R., Hofkosh, D., Barone, M. A., & Holmboe, E. S. (2017). Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Academic Medicine*, *92*(3), 324–330. https://doi.org/10.1097/ACM.0000000000001141
12. Cavallaro, S. C., Fishman, M. D., Dorney, K., Hirsch, A. W., Maheshwari, A., Nagler, J., Pusic, M. V., Schutzman, S., Westervelt, M., & Miller, K. A. (2026). Fostering conversation and co-regulation: how medical students experience co-creating narrative feedback. *Academic Medicine*, wvag202. https://doi.org/10.1093/acamed/wvag202
13. Ceccim, R. B., & Feuerwerker, L. C. M. (2004). O quadrilátero da formação para a área da saúde: ensino, gestão, atenção e controle social. *Physis: Revista de Saúde Coletiva*, *14*(1), 41–65. https://doi.org/10.1590/S0103-73312004000100004
14. Charmaz, K. (2009). *A construção da teoria fundamentada: guia prático para análise qualitativa*. Artmed.
15. Charmaz, K. (2017). The power of constructivist grounded theory for critical inquiry. *Qualitative Inquiry*, *23*(1), 34–45. https://doi.org/10.1177/1077800416657105
16. Costa, L. B., Esteche, F. F., Filho, A., Bomfim, R. F., A. L. B., & Ribeiro, M. T. A. M. (2018). Competências e atividades profissionais confiáveis: novos paradigmas na elaboração de uma matriz curricular para residência em Medicina de Família e Comunidade. *Revista Brasileira de Medicina de Família e Comunidade*, *13*(40), 1–11. https://doi.org/10.5712/rbmfc13(40)1632
17. Danilovich, N., Kitto, S., Price, D. W., Campbell, C., Hodgson, A., & Hendry, P. (2021). Implementing competency-based medical education in family medicine: a narrative review of current trends in assessment. *Family Medicine*, *53*(1), 9–22. https://doi.org/10.22454/FamMed.2021.453158
18. Englander, R., Frank, J. R., Carraccio, C., Sherbino, J., Ross, S., & Snell, L.,ICBME Collaborators (2017). Toward a shared language for competency-based medical education. *Medical Teacher*, *39*(6), 582–587. https://doi.org/10.1080/0142159X.2017.1315066

19. Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal Of The American Medical Association*, *287*(2), 226–235. https://doi.org/10.1001/jama.287.2.226
20. Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in Health Sciences Education*, *16*(3), 311–329. https://doi.org/10.1007/s10459-010-9263-2
21. Frank, J. R., Snell, L. S., Cate, T., Holmboe, O., Carraccio, E. S., Swing, C., Harris, S. R., Glasgow, P., Campbell, N., Dath, C., Harden, D., Iobst, R. M., Long, W. F., Mungroo, D. M., Richardson, R., Sherbino, D. L., Silver, J., Taber, I., Talbot, S., M., & Harris, K. A. (2010). Competency-based medical education: theory to practice. *Medical Teacher*, *32*(8), 638–645. https://doi.org/10.3109/0142159X.2010.501190
22. Fraundorf, S. H., Caddick, Z. A., Nokes-Malach, T. J., & Rottman, B. M. (2023). Cognitive perspectives on maintaining physicians' medical expertise: III. Strengths and weaknesses of self-assessment. *Cognitive Research: Principles and Implications*, *8*, 58. https://doi.org/10.1186/s41235-023-00511-z
23. Fuentes-Cimma, J., Sluijsmans, D., Riquelme, A., Villagrán, I., Isbej, L., Olivares-Labbe, M. T., & Heeneman, S. (2024). Designing feedback processes in the workplace-based learning of undergraduate health professions education: a scoping review. *BMC Medical Education*, *24*, 440. https://doi.org/10.1186/s12909-024-05439-6
24. Gantayet-Mathur, A., Chan, K., & Kalluri, M. (2022). Patient-centered care and interprofessional collaboration in medical resident education: where we stand and where we need to go. *Humanities and Social Sciences Communications*, *9*, 206. https://doi.org/10.1057/s41599-022-01221-5
25. Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the black box differently: assessor cognition from three research perspectives. *Medical Education*, *48*(11), 1055–1068. https://doi.org/10.1111/medu.12546
26. Govaerts, M., & Van der Vleuten, C. P. M. (2013). Validity in work-based assessment: expanding our horizons. *Medical Education*, *47*(12), 1164–1174. https://doi.org/10.1111/medu.12289
27. Heim, M., Schulz, C. M., Schneider, F., Berberat, P. O., Gartmeier, M., & Schick, K. (2023). Measuring informal workplace learning outcomes in residency training: a validation study. *BMC Medical Education*, *23*, 549. https://doi.org/10.1186/s12909-023-04529-1
28. Hempel, A., Boyle, J., de Bruin, A., & Ginsburg, S. (2026). Just a checkbox: growth mindset and feedback in resident experiences with EPA assessments. *Medical Education*, 1–9. https://doi.org/10.1111/medu.70234
29. Johnson, C. E., Keating, J. L., Farlie, M. K., Kent, F., Leech, M., & Molloy, E. K. (2019). Educators' behaviours during feedback in authentic clinical practice settings: an observational study and systematic analysis. *BMC Medical Education*, *19*, 129. https://doi.org/10.1186/s12909-019-1524-z
30. King, R., Downer, T., Lord, B., Flanagan, B., & Oprescu, F. (2025). A practical example of how to apply constructivist grounded theory methodology: exploring patient experiences during paramedic-led healthcare. *Research in Nursing & Health*, *48*(4), 508–521. https://doi.org/10.1002/nur.22468
31. Lerch, S. P., Huwendiek, S., Nendaz, M. R., Klöppel, S., & Pinilla, S. (2025). Perceptions of ad hoc entrustment, the need for supervision, and coping strategies in clinical residents: a qualitative study. *Academic Medicine*, *100*(7), 844–851. https://doi.org/10.1097/ACM.0000000000006019
32. Lingard, L. (2009). What we see and don't see when we look at competence: notes on a god term. *Advances in Health Sciences Education*, *14*(5), 625–628. https://doi.org/10.1007/s10459-009-9206-y
33. Machado, L. B. M., Marques, C. C., Rodrigues, L. C. M., Sperling, S., Machado, N. C., Gusso, G. D. F., & Valladão Júnior, J. B. R. (2018). O currículo de competências do programa de residência em Medicina de Família e Comunidade da Faculdade de Medicina da Universidade de São Paulo. *Revista Brasileira de Medicina de Família e Comunidade*, *13*(40), 1–16. https://doi.org/10.5712/rbmfc13(40)1602
34. Matuda, C. G., Pinto, N. R. S., Martins, C. L., & Frazão, P. (2015). Colaboração interprofissional na Estratégia Saúde da Família: implicações para a produção do cuidado e a gestão do trabalho. *Ciência & Saúde Coletiva*, *20*(8), 2511–2521.

https://doi.org/10.1590/1413-81232015208.11652014

35. Mukhia, S., & Tenzin, K. (2025). Training academic staff for effective feedback in workplace-based assessment: a study in Bhutan. *BMC Medical Education*, *25*, 748. https://doi.org/10.1186/s12909-025-07314-4
36. Sargeant, J., Lockyer, J., Mann, K., Holmboe, E., Silver, I., Armson, H., Driessen, E. W., MacLeod, T., Yen, W., Ross, K. M., & Power, M. E. (2015). Facilitated reflective performance feedback: developing an evidence- and theory-based model that builds relationship, explores reactions and content, and coaches for performance change (R2C2). *Academic Medicine*, *90*(12), 1698–1706. https://doi.org/10.1097/ACM.0000000000000809
37. Sargeant, J., Lockyer, J. M., Mann, K., Armson, H., Warren, A. E., Zetkulic, M., Soklaridis, S., Könings, K. D., Ross, K. M., Silver, I., Holmboe, E. S., Shearer, C. L., & Boudreau, M. (2018). The R2C2 model in residency education: how does it foster coaching and promote feedback use? *Academic Medicine*, *93*(7), 1055–1063. https://doi.org/10.1097/ACM.0000000000002131
38. Schultz, K., McGregor, T., Pincock, R., Nichols, K., Jain, S., & Pariag, J. (2022). Discrepancies between preceptor and resident performance assessment: using an electronic formative assessment tool to improve residents' self-assessment skills. *Academic Medicine*, *97*(5), 669–673. https://doi.org/10.1097/ACM.0000000000004154
39. Schuwirth, L. W. T., & Van der Vleuten, C. P. M. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical Teacher*, *33*(6), 478–485. https://doi.org/10.3109/0142159X.2011.565828
40. Separavich, M. A., & Couto, M. T. (2021). Programa Mais Médicos: revisão crítica da implementação sob a perspectiva do acesso e universalização da atenção à saúde. *Ciência & Saúde Coletiva*, *26*(Suppl 2), 3435–3446. https://doi.org/10.1590/1413-81232021269.2.04572020
41. Shafian, S., Ilaghi, M., Shahsavani, Y., Okhovati, M., Soltanizadeh, A., Aflatoonian, S., & Karamoozian, A. (2024). The feedback dilemma in medical education: insights from medical residents' perspectives. *BMC Medical Education*, *24*, 424. https://doi.org/10.1186/s12909-024-05398-y
42. Soares, R. S., Oliveira, F. P., Neto, M., Barrêto, A. J., Carvalho, D. S., Sampaio, A. L. B., J., & Figueiredo, A. M. (2018). Residência em Medicina de Família e Comunidade: construindo redes de aprendizagens no SUS. *Revista Brasileira de Medicina de Família e Comunidade*, *13*(40), 1–8. https://doi.org/10.5712/rbmfc13(40)1629
43. Souza, E. F. D. (2024). *Inventário de competências: lacunas e potencialidades formativas nas residências em Medicina de Família e Comunidade e multiprofissional em Saúde da Família no estado da Paraíba* [Dissertação, Universidade Estadual da Paraíba]. https://tede.bc.uepb.edu.br/jspui/handle/tede/5140
44. Souza, E. F. D., Machado, A. P. R., Oliveira, J. D. L., Oliveira, J. O. D., Nóbrega, L. G. S., Toscano, M. B. R., Silva, D. M., Olinda, R. A., Nóbrega, R. V., Matheson, D., & Santos, S. (2025). Inventory of self-assessed competences: a tool for identifying training strengths and weaknesses in medical and multiprofessional residency programmes for primary health care. *BMC Medical Education*, *25*, 944. https://doi.org/10.1186/s12909-025-07521-z
45. Švab, I., & Cerovečki, V. (2024). Person-centred, a core concept of family medicine. *European Journal of General Practice*, *30*(1), 2393860. https://doi.org/10.1080/13814788.2024.2393860
46. Telio, S., Ajjawi, R., & Regehr, G. (2015). The educational alliance as a framework for reconceptualizing feedback in medical education. *Academic Medicine*, *90*(5), 609–614. https://doi.org/10.1097/ACM.0000000000000560
47. Ten Cate, O. (2017). Competency-based postgraduate medical education: past, present and future. *GMS Journal for Medical Education*, *34*(5), Doc69. https://doi.org/10.3205/zma001146
48. Ten Cate, O., & Scheele, F. (2007). Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Academic Medicine*, *82*(6), 542–547. https://doi.org/10.1097/ACM.0b013e31805559c7
49. Thakur, A., Darani, S., Kalocsai, C., Silver, I., Sockalingam, S., & Soklaridis, S. (2023). Implementation of structured feedback in a psychiatry residency program in Canada: a qualitative analysis study. *Frontiers in Psychiatry*, *14*, 1276985. https://doi.org/10.3389/fpsyt.2023.1276985

50. Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042
51. Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, *34*(3), 205–214. https://doi.org/10.3109/0142159X.2012.652239
52. Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education*, *53*(1), 76–85. https://doi.org/10.1111/medu.13645
53. Zafošnik, U., Cerovečki, V., Stojnić, N., Poženel Belec, A., & Klemenc-Ketiš, Z. (2024). Developing a competency framework for training with simulations in healthcare: a qualitative study. *BMC Medical Education*, *24*, 180. https://doi.org/10.1186/s12909-024-05139-1

Figures

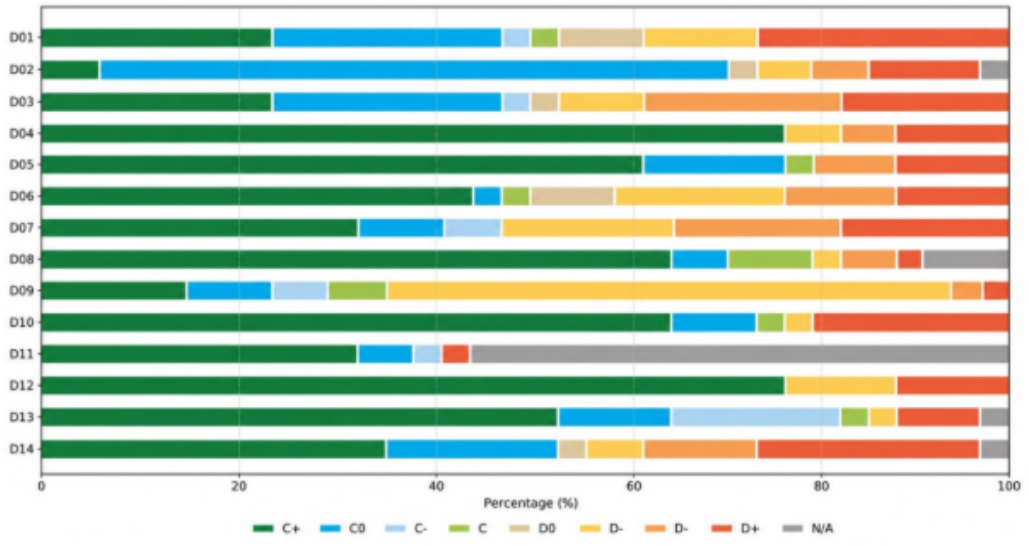


Figure 1

Distribution of evaluative (mis)alignment categories across the participating dyads. Categories are defined as follows: N/C, not classified; C0, agreement without justification; CI, agreement by inference; C−, agreement with different justifications; C+, agreement with aligned justifications; D0, disagreement without justification; DI, disagreement by inference; D−, preceptor underestimation; and D+, preceptor overestimation.



Figure 2

Pedagogical model for structured feedback — S-SKILLS-F.

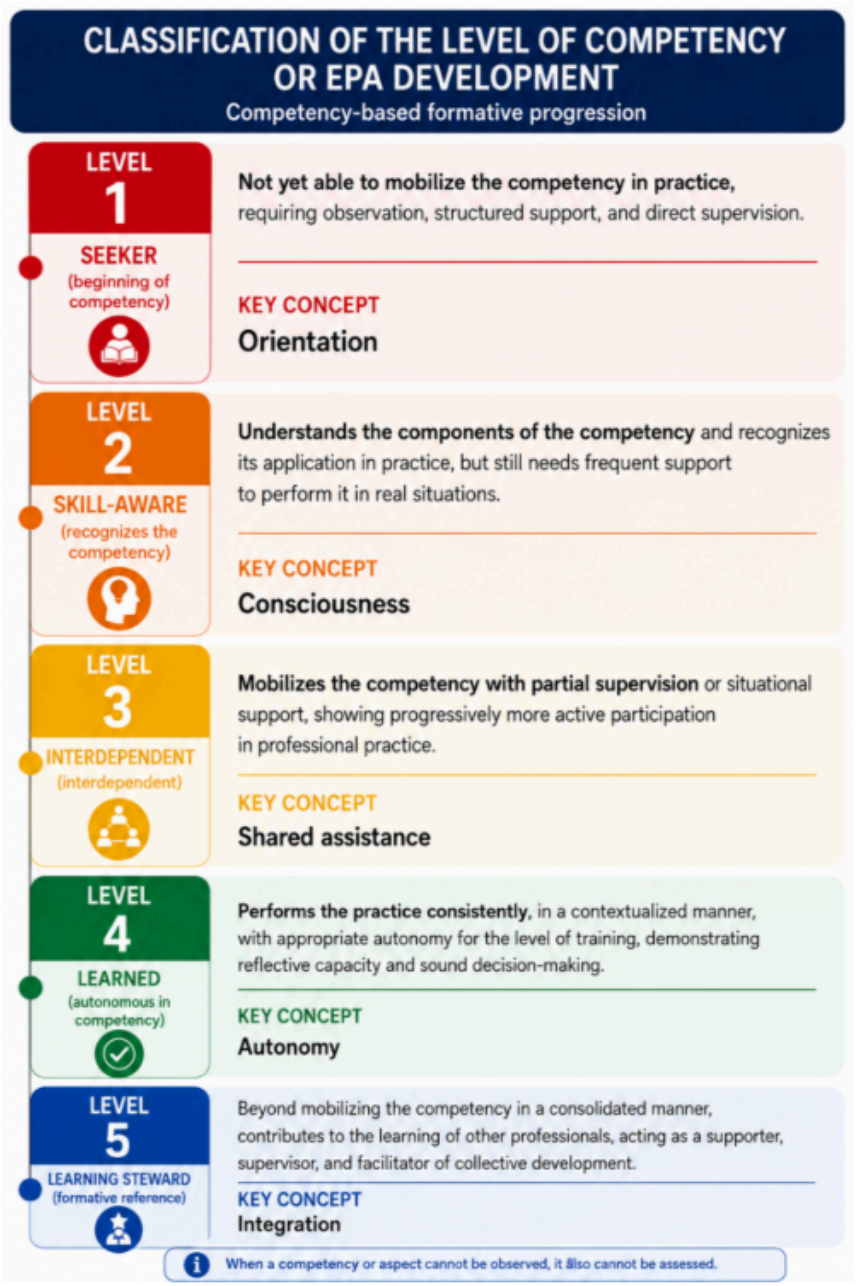


Figure 3

Classification of competency development levels in S-SKILLS-F.