



OPEN ACCESS

EDITED BY

Evely Boruchovitch,
State University of Campinas, Brazil

REVIEWED BY

Yafes Can,
Anadolu University, Türkiye
Carolina Tapia Cortés,
University of Monterrey, Mexico

*CORRESPONDENCE

Laura Schwitter
✉ laura.schwitter@phbern.ch

RECEIVED 05 January 2026

REVISED 11 May 2026

ACCEPTED 15 June 2026

PUBLISHED 03 July 2026

CITATION

Schwitter L, Eckstein B, Wettstein A and Fauth B (2026) The multi-perspective assessment of pedagogical interaction quality as a tool for teaching development: insights from the INTAKT pilot study.
Front. Educ. 11:1781315.
doi: 10.3389/feduc.2026.1781315

COPYRIGHT

© 2026 Schwitter, Eckstein, Wettstein and Fauth. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The multi-perspective assessment of pedagogical interaction quality as a tool for teaching development: insights from the INTAKT pilot study

Laura Schwitter^{1*}, Boris Eckstein², Alexander Wettstein¹ and Benjamin Fauth³

¹Institute for Research and Development, University of Teacher Education Bern, Bern, Switzerland,

²Institute for Research, Development and Evaluation, University of Teacher Education Zurich, Zurich, Switzerland, ³University of Education Freiburg, Freiburg, Germany

Effective pedagogical interactions are an important prerequisite for the cognitive, emotional, and motivational development of students and the health of teachers. Yet the dynamics of pedagogical interactions have not yet been sufficiently conceptualized or adequately researched from multiple perspectives. The INTAKT study addresses these gaps by investigating teachers' and students' behaviors and experiences in secondary school classes—from their subjective perspectives (insider view) and by classroom observations (external view). In this paper, we present the INTAKT pilot study, which examined the feasibility of this multi-perspective approach in 16 classes ($n = 209$ students). Descriptive analyses yielded insightful results on the (perceived) quality of 18 interaction features from the distinct perspectives in each participating class, e.g., visualized by radar charts. Effect sizes suggest substantial differences across perspectives, with the largest and most consistent discrepancy observed for cognitive activation ($d' > 1.0$). We exemplify these results with two contrast groups built on observer ratings (low vs. high quality). A comparison of these two groups indicates that the perspective differences were larger in the low-quality group. In addition, we present how we translated the findings into personalized feedback reports for the participating teachers. As an outlook, we document preliminary psychometric diagnostics. In the discussion, we argue that the INTAKT approach is feasible and yields meaningful findings with high face validity. As for the feedback reports, we discuss their potential to support teachers in reflecting on and developing their instructional practice. Beyond that, we critically reflect on the pilot study's sample quality (small number of classes, convenience sample) and outline the main study's strategy to further analyze the scales' psychometric properties and the generalizability of the observed effects.

KEYWORDS

classroom climate, classroom feedback, classroom management, multi-perspective assessment, pedagogical interaction, secondary education, teaching and learning

1 Introduction

Pedagogical interactions refer to the reciprocally related experiences and behaviors of interaction partners in the classroom and form the framework for institutional teaching and learning processes (Perrez et al., 2006; Wubbels et al., 2023). They are crucial for the cognitive, emotional, and motivational development of students (Pianta et al., 2003; Scherzinger and Wettstein, 2022). Recent longitudinal evidence further highlights the central role of teacher-student relationships for student engagement and achievement across developmental stages (Di Lisio et al., 2025). In addition, pedagogical interactions impact the professional and mental well-being of teachers (Kokkinos, 2007).

Despite their significance, the dynamics of pedagogical interactions in the classroom have not yet been sufficiently conceptualized in theory or adequately researched in empirical studies (Eckstein and Wettstein, 2024; Fauth et al., 2020). Existing instruments frequently consider only one perspective—that of the teacher, the student or the observer—and measure exclusively either observable behavior or subjective experience. Research on teaching quality measurement has shown that ratings from students, teachers, and external observers frequently exhibit low agreement (Wang et al., 2020), suggesting that such single-perspective designs capture only partial aspects of instructional quality and may contribute to the small and inconsistent effects recently highlighted in meta-analytic and review work in this field (Charalambous and Praetorius, 2025). This reinforces the case for multi-perspective approaches that capture the complementary information held by different actors. Even recent scale development efforts, while advancing the measurement of teacher-student interaction, illustrate this limitation by predominantly relying on single-perspective self-reports and not systematically integrating multiple perspectives or distinguishing between behavioral and experiential components (e.g., Can and Guven, 2025).

Classroom research has established a theoretically and empirically grounded framework for systematically assessing teaching quality, in the German tradition widely known as the model of three basic dimensions: classroom management, emotional support, and cognitive activation (Bleck and Lipowsky, 2019; Klieme et al., 2001; Praetorius et al., 2020). However, this framework—similar to the CLASS approach (Pianta, 2016; Pianta and Hamre, 2009)—focuses primarily on teacher behavior as observed from an external perspective and does not systematically distinguish between the contributions of teachers and students, nor between their observable behavior and subjective experiences. A lack of evidence-based knowledge exists regarding the extent to which (i) the objectively observable behavior of teachers and students or (ii) their subjective experiences determine the quality of pedagogical interactions. For instance, a classroom disturbance may be perceived as a minor issue by the students, whereas the teacher may experience it as much more stressful (Eckstein and Wettstein, 2024). These gaps point to the need for a multimethod, multi-perspective approach that can capture both behavioral and experiential aspects of interaction quality—and that is feasible in classroom research and meaningful for use in teaching development.

1.1 Differences in perspectives

Previous studies on teaching quality reveal a tendency for low relative agreement between student and teacher assessments of teaching quality (Clausen, 2002; Desimone, 2009; Fauth et al., 2014, 2020; Kunter and Baumert, 2006; Scherzinger and Wettstein, 2019; Wagner et al., 2016; Wettstein et al., 2018). These discrepancies can primarily be explained by the different roles and perspectives of teachers, students, and—more recently—external observers in the classroom context.

Teachers are usually expected to assess the quality of interaction in their classrooms with reasonable accuracy based on their professional expertise, particularly regarding structural features or instructional settings (Fauth et al., 2020). However, empirical studies indicate that they tend to evaluate their own classroom management as more consistent and effective than their students or external observers do (Kunter and Baumert, 2006). This may be due to the fact that teachers must simultaneously manage and monitor several teaching processes, which makes it challenging to adequately perceive single interactions (Helmke and Lenske, 2013). Furthermore, teachers may present their own behavior in a socially desirable light or overemphasize external circumstances (Kunter and Baumert, 2006; Weiner, 1986; Wubbels et al., 1992).

In contrast, students encounter teaching from a less involved perspective—that is, with less immediate pressure to act—enabling them to observe particular interactions in a more objective manner compared to that of the teacher (Wettstein, 2013). Yet, student perceptions vary considerably, as ratings within the same class frequently reveal minimal consensus (Wagner et al., 2016). This suggests that students perceive lessons differently, presumably due to individual expectancies or the quality of their relationship with the teacher (Fauth et al., 2014). Moreover, positive teacher-student relationships can lead to halo effects, whereby students rate other aspects of the lesson, such as the clarity of explanations, at a higher level (Fauth et al., 2020).

Finally, external observers are expected to provide an intersubjectively transparent assessment of teaching processes, as they evaluate from an unbiased external standpoint (Helmke, 2009). This perspective is further enhanced through standardization, including video-based observation and specific coding schemes (Helmke, 2009; Praetorius et al., 2012; Scherzinger and Wettstein, 2019; Wettstein et al., 2018). Video recordings of a limited segment of a teaching scenario may, however, compromise ecological validity, as they are constrained to a specific time frame and cannot provide feedback on the teaching and relationship dynamics beyond this segment (Praetorius et al., 2017).

Helmke and Lenske (2013) therefore call for explicit consideration of different perspectives in classroom diagnostics and a stronger focus on differences in perspective. This call is central to the present study: by systematically combining teacher, student, and observer perspectives, a multi-perspective design can capitalize on the complementary strengths of each informant while mitigating their individual limitations—provided such a design is feasible in classroom research.

1.2 INTAKT

Building on this, the overarching aim of the INTAKT research project is to develop and test a multi-perspective, multimethod

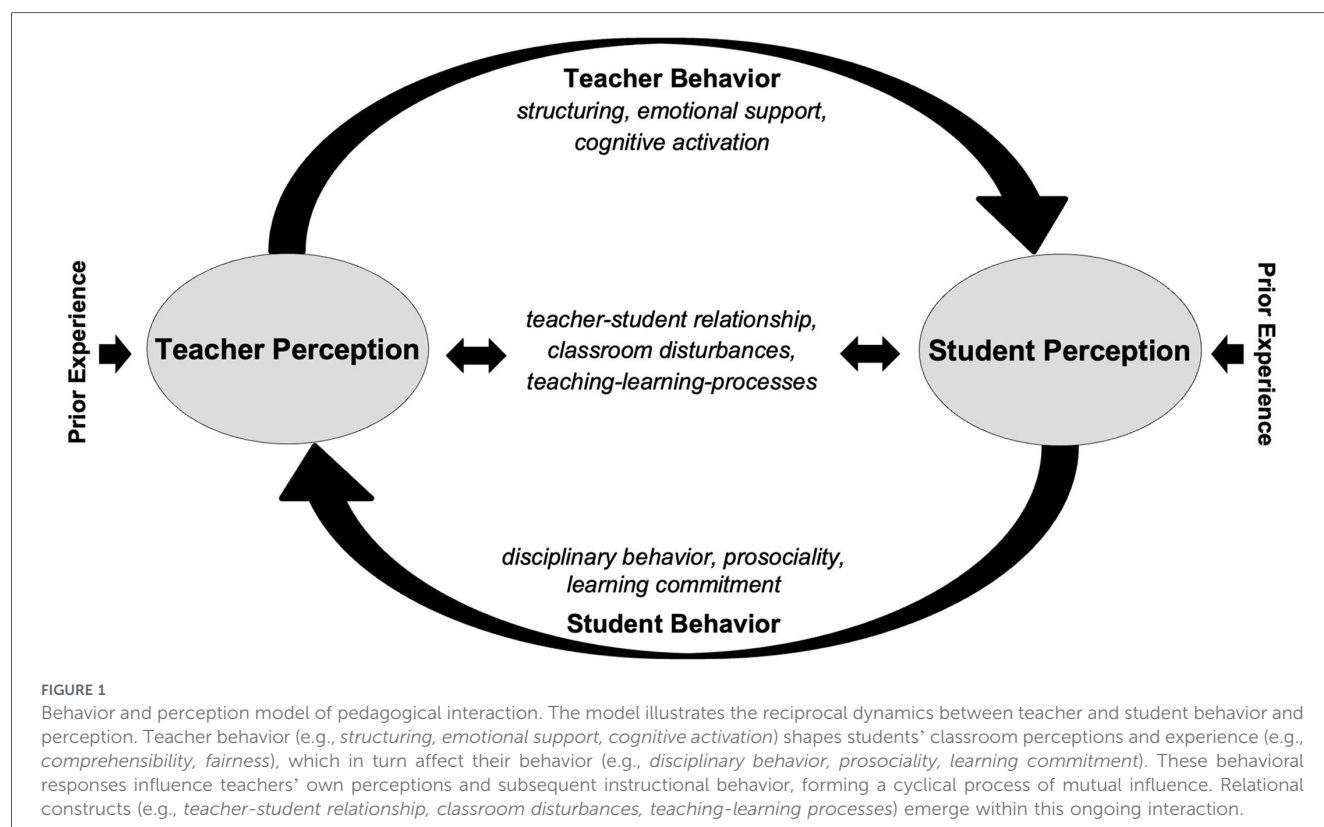
approach to assessing the quality of pedagogical interactions, integrating both observable behaviors and the subjective experience of teachers and students. In this article, we present a pilot study that examines the feasibility of this approach. Specifically, the INTAKT pilot study explores perspective differences between teachers, students, and observers across constructs, as well as how the findings can be presented to teachers in an evidence-based and practically meaningful way to support instructional practice. The theoretical foundation of our approach is a process-oriented understanding of pedagogical interactions that links different types of features (Eckstein and Wettstein, 2024; Perrez et al., 2006; Thiel, 2016): (1) *Behavior* refers to all activities of teachers and students, such as providing and following instructions or asking questions (Eckstein and Wettstein, 2024; Perrez et al., 2006; Pianta et al., 2003). (2) *Subjective experience* refers to how teachers and students perceive, interpret, and appraise the behavior of their interaction partners and general classroom features, such as the comprehensibility, learning conditions, or interpersonal climate (Eckstein and Wettstein, 2024; Schweer et al., 2017; Thies, 2017). *Behavior* and *subjective experience* are reciprocally related: teacher behavior affects students' subjective experience, which then shapes their behavioral response. The students' behavior, in turn, also influences teachers' subjective experience and subsequent activities, underlining the cyclical nature of pedagogical interactions.

Pedagogical interactions are dynamic and highly context dependent. Figure 1 illustrates the main postulated mechanisms of activities in the *Behavior and Perception Model of Pedagogical Interaction*, a theoretical process model adapted from Eckstein and Wettstein (2024) and based on preliminary work by Wettstein (2012) and Eckstein (2018). The model conceptualizes

the interplay between behavior and subjective experience as a circular, transactional process that unfolds over successive stages of interaction. Teachers' behavior—such as *structuring*, providing *emotional support*, and promoting *cognitive activation*—shapes how students experience the classroom, for instance in terms of comprehensibility and perceived fairness. Students' experiences, in turn, affect their subsequent behavior, including *disciplinary behavior*, prosocial behavior, and *learning commitment*. These behavioral patterns then influence teachers' experiences—for example, feelings of satisfaction or irritation—which subsequently guide their instructional behavior in the next interaction phase. These interactions do not occur in isolation but are embedded in ongoing feedback loops: the behavior and experience of one interactional sequence influence subsequent ones, resulting in a dynamic, cyclical process (Brunner and Noack, 2010). Moreover, this process is shaped by individual characteristics (Schweer et al., 2017; Thies, 2017; Werth et al., 2020), such as the teacher's patience and professional expertise or the student's prior knowledge and learning commitment, as well as by the contextual conditions of the classroom, such as the quality of relationships (Thiel, 2016), and the (extra)curricular context (Doyle, 2006).

Within this reciprocal system, relational phenomena—such as the teacher-student relationship, teaching-learning processes, and classroom disturbances—emerge as shared interactional phenomena. The overall quality of pedagogical interaction thus develops through the continuous interplay between teacher and student behavior, their experience, and the relational processes that connect them.

Depending on how these processes unfold, pedagogical interactions may be functional or dysfunctional. Functional



interactions are associated with enhanced teaching and learning processes, strengthened positive social relationships, and creation of a favorable and motivating learning environment. Dysfunctional interactions, on the other hand, may hinder the teaching and learning processes, burden the participants' emotional and motivational experiences, and are often associated with classroom disturbances (Brunner and Noack, 2010; Pianta et al., 2012; Scherzinger and Wettstein, 2022).

1.3 Three dimensions of interaction quality: the INTAKT taxonomy

The INTAKT taxonomy systematizes quality features of pedagogical interactions along three dimensions of distinction derived from established frameworks in teaching quality research (Fend, 1980; Klieme et al., 2001; Pianta and Hamre, 2009; Praetorius et al., 2020) and our process model introduced above (Eckstein and Wettstein, 2024). By integrating these dimensions of distinction, the taxonomy provides the conceptual basis for operationalizing pedagogical interaction quality across perspectives. For the present pilot study, 18 constructs were operationalized from this taxonomy (See Table 1). The following subsections address the three dimensions of distinction individually. Subsequently, we elaborate on how the taxonomy integrates them into a coherent cross-classification system.

1. Actor. Rooted in co-constructivist theory, supply-use models (e.g., Fend, 1980; Vieluf et al., 2020), describe teaching and learning processes as interactions between

the learning opportunities provided by the teacher and their use by the students. According to this understanding, features of pedagogical interaction quality can be attributed to either (i) teachers or (ii) students. *Teachers* influence the quality of interaction, for example, by structuring and guiding the lesson: they create the conditions for effective teaching and learning processes and positive socio-emotional dynamics. *Students* contribute to the interaction quality not only through their behavior—such as their engagement in learning or their social and disciplinary behavior—but also through their subjective experience. Emotional reactions, such as feelings of fairness or support, are also part of the use of the learning opportunity and have a significant impact on the quality of interaction.

2. Domain. Informed by empirical evidence, frameworks of teaching quality research categorize features into thematic domains. The framework of the three basic dimensions of teaching quality refers to these domains as classroom management, emotional support, and cognitive activation (Bleck and Lipowsky, 2019; Klieme et al., 2001; Praetorius et al., 2020). In our approach, drawing on the CLASS framework (Pianta and Hamre, 2009), the labels for these domains have been adapted to equally reflect the perspectives of teachers and students, as well as their behaviors and experiences. Accordingly, the teaching domain dimension distinguishes three central areas that are particularly relevant for analyzing the quality of pedagogical interactions: (i) *organization*, (ii) *climate*, and (iii) *knowledge co-construction*.

- (i) *Organization* encompasses, on the part of the teacher, classroom management and the assurance of smooth instructional flow, including the establishment of clear rules, routines, and monitoring of classroom processes. On behalf of the students, this domain encompasses disciplinary behavior and academic engagement. A key characteristic of functional pedagogical interactions in the organizational domain is clear and consistent classroom management, which supports learning processes through rules, routines, and transparent behavioral guidelines (Hattie, 2012; Seidel, 2015). Disorganized classroom management, marked by unclear expectations and insufficient behavioral regulation, often results in classroom disturbances (Klieme, 2019; Wettstein and Scherzinger, 2022). Students' disciplinary behavior also plays a crucial role in determining instructional quality. Functional classroom interactions are characterized by compliant behavior and active participation, which are considered beneficial to learning. In contrast, indisciplined student behavior can disrupt instruction and negatively affect both teachers and classmates (Arbuckle and Little, 2004; Beaman et al., 2007). However, student indiscipline is not experienced uniformly. Among teachers, for example, those with high self-efficacy tend to

TABLE 1 Overview of constructs of pedagogical interaction quality.

Teacher behavior	Relational/General experience	Student behavior
Organization		
Structuring	Classroom Disturbances	Engagement
Monitoring	Working Atmosphere	Disciplinary Behavior
Climate		
Emotional support	Teacher-Student Relationship	Prosocial Behavior Students Teacher
Teacher aggression	Well-being	Student Aggression
Knowledge co-construction		
Cognitive activation	Teaching-Learning Processes	Learning Commitment
Instructional support	Intelligibility	

Table presents the 18 constructs of pedagogical interaction quality derived from the INTAKT taxonomy. The table differentiates three teaching domains (*organization*, *climate*, and *knowledge co-construction*; vertical structure), two actors (*teacher* and *student*; horizontal structure), and three actor-specific feature types (*behavior*, *perception of behavior*, and *general experience*). The constructs shown here reflect mainly the *behavioral* and *experiential* aspects of these processes, as the *perception of behavior* is captured indirectly through the different perspectives. The middle column includes *relational constructs* (e.g., teacher-student relationship) and *general experiences* (e.g., *atmosphere*, *well-being*, *intelligibility*) that cannot be clearly assigned to either actor.

interpret disturbances with greater equanimity, whereas teachers with low self-efficacy are more likely to experience them as stressful and disruptive. From the students' perspective, clear rules and expectations are typically experienced as structuring and supportive (Dicke et al., 2014; Jang et al., 2010).

- (ii) *Climate* refers to the emotional and motivational support provided by the teacher as well as the social relationships in the classroom. It is reflected in the warmth, appreciation, and mutual support that teachers and students express toward one another. Classroom climate is co-constructed by both the teacher's and the students' behavior. In particular, emotional support and clear communication from the teacher, as well as appreciative teacher-student relationships, foster a positive social climate (Schwarzwälder and Lohse-Bossenz, 2021). In addition, cooperative learning methods and proactive conflict prevention strengthen social cohesion (Lanphen and Wiedenbauer, 2021). Students also contribute particularly through prosocial behavior, which actively supports a positive and cohesive classroom climate (Siegler et al., 2016). Beyond the benefits for students—such as greater social cohesion, mutual support, and a more motivating learning environment—teachers likewise profit from a positive classroom climate, which has been shown to reduce stress and increase job satisfaction (Grewé, 2016).
- (iii) *Knowledge co-construction* focuses on cognitively activating instruction that promotes understanding, independent thinking, and deep processing of content, alongside sustained student effort. Functionally, it is characterized by challenging tasks and a productive error culture (Klieme, 2019). Teachers contribute by preparing content and developing questions and tasks that stimulate higher-order thinking (Klieme, 2019). Students contribute when they maintain focus, engage deeply with content, and connect ideas (Dweck, 2000). Instructional support that is conducive to learning, however, requires an accurate assessment of students' current understanding (Kunter et al., 2020; Timmermans et al., 2021).

Feature type. Drawing on interactionist theory (e.g., Perrez et al., 2006; Wubbels et al. 2015), and the psychology of social perception (e.g., Hall et al., 2016), we conceptualize pedagogical interactions as transactional cycles of the actors' mutual behaviors and experiences (see Figure 1). The constituents of these procedures can be further differentiated into four feature types. These comprise, first, three actor-specific core features: (i) behavior, (ii) behavior perception, and (iii) general experience. Second, they comprise (iv) relational features of the overall interaction system. (i) *Behavior* refers to the activities of the actors during the interaction, such as a teacher providing instructions or students listening actively. (ii) *Behavior*

perception refers to how these activities are noticed and processed cognitively, emotionally, and motivationally by the interaction partners. This is relevant because the subjective perception of a behavior does not necessarily correspond to the actor's initial intention. (iii) *General experience* refers to the actors' global impressions, emotions, and motivational states that emerge during the interaction, including aspects such as the social climate and the actors' wellbeing. This goes beyond perceptual processing of individual behaviors, as it is shaped by broader emotional and motivational factors. (iv) *Relational features* of pedagogical interactions emerge from the interplay of teachers' and students' mutual contributions; they characterize the transactional and reciprocal nature of the overall interaction system, —for example, *classroom disturbances* or the *teacher-student relationship*. Since both teachers and students contribute to relational features, they cannot be attributed exclusively to either.

The INTAKT taxonomy integrates these three dimensions of distinction into a cross-classification system comprising 21 conceptual units of pedagogical interaction quality: 2 actors (teacher, student) $\times$ 3 domains (organization, climate, knowledge co-construction) $\times$ 3 core features (behavior, behavior perception, general experience) + 3 relational features. It should be noted that the taxonomy is intended as an analytical guideline for operationalization. That is, its 21 units represent umbrella categories rather than first-order constructs. For the present pilot study's operationalization, we further differentiated certain aspects, e.g., teacher behavior within the organization domain (not only one, but two features). By contrast, it was not possible to materialize the analytical distinction between behavior and behavior perception. That is, our empirical measures of behavior are, in fact, measures of perceived behavior, e.g., student-perceived teacher behavior. Still, the analytical distinction remains meaningful, since we plan to approximate actors' "true" behaviors by integrating multiple measures of perceived behavior from different informants in later in-depth analyses. Eventually, 18 constructs have been operationalized and explored in the present pilot study.

Conceptually, the INTAKT taxonomy frames pedagogical interaction quality as a multidimensional construct. While overarching dimensions are theoretically clearly distinguishable, they are also assumed to be empirically interrelated. Accordingly, each of the 18 constructs is expected to be unidimensional within a given perspective, with teachers, students, and observers rating equivalent content in perspective-adapted wording. However, the theoretical classification into the three teaching domains (organization, climate, knowledge co-construction) is not intended to be empirically tested as a higher-order factor structure at this stage. A more comprehensive psychometric examination—including measurement invariance across perspectives—requires larger samples and is therefore reserved for the ongoing INTAKT main study with a sample of 120 classes ($n \approx 1,870$ students). Initial analyses of the first wave data indicate strong scale properties; full psychometric validation results will be reported in a forthcoming companion paper.

In the present pilot study, the feasibility of this operationalization was examined by applying it in two ways to test whether it yields meaningful and interpretable data in

classroom research: first, at the classroom level to generate evidence-based feedback reports for teachers to support their teaching development (Brühwiler and Leutwyler, 2020). The ratings from the different perspectives can be presented in visual form, enabling similarities and divergences in perception to be identified at a glance. To ensure that these class-level feedback reports are both interpretable and supportive of teacher reflection and instructional development, the INTAKT approach relies on absolute perspective differences (d) and standardized effect sizes (e.g., Cohen's d) rather than more abstract statistical indicators such as correlation coefficients. This choice reflects both practical and methodological considerations: effect sizes are typically more accessible and intuitively understandable for teachers (Valentine et al., 2015), especially when presented visually. In addition, they allow for meaningful interpretation even in the context of small sample sizes, where traditional significance testing would lack power and relevance (Lind, 2012). Alternatively, class-level results can be compared with values from the overall sample to situate them within a broader context.

Beyond that, effect sizes at the sample level provide preliminary insights into patterns and variation of the constructs—including cross-perspective differences—when generalizability is limited due to small sample size. In small samples, aggregating classes with comparatively high or low levels of interaction quality into two theoretically informed contrast groups can provide a pragmatic and empirically grounded approximation of the range of interaction quality. Contrasting these two groups enables an anonymized and theoretically grounded representation of variability in pedagogical interaction quality and illustrates the added value of multi-perspective diagnostics even in small and heterogeneous samples. Moreover, sample-level analyses in classroom research often face methodological constraints such as low case numbers per unit (e.g., class or teacher), which limit the use of inferential statistics or correlational designs. This is particularly relevant in multi-perspective studies, where differentiated constructs are assessed across actors, but each perspective may only yield one data point per class. Against this background, focusing on descriptive measures and effect-size-based group comparisons offers a robust and practically grounded approach to mapping variation in pedagogical interaction quality.

1.4 Research questions

Previous research approaches on pedagogical interactions have provided valuable insights into teachers' and students' behavior, differences in perspectives, and their relevance for teaching (Klieme et al., 2001; Pianta and Hamre, 2009; Praetorius et al., 2020). However, empirical research so far has often focused on analyzing the observable behavior of teachers and students, while subjective experiences, such as perception, interpretation, and evaluation of interactions, have been less systematically considered. For instance, classroom discipline can be understood not only in terms of objectively established rules but also through the teacher's perception of student discipline—and the impact of these (more or less accurate) perceptions on the subsequent course of interaction. A key area of interest is the

interplay between behavior and experience. In addition, methodological challenges have emerged in previous research: differences in perspective between teachers and students may have been reinforced in some cases by imprecise operationalizations of pedagogical interaction quality (Clausen, 2002). This requires further development of existing measurement approaches to enable a more precise assessment of the quality of pedagogical interaction. Along with the question of how to assess the quality of pedagogical interactions from multiple perspectives, it is important to consider how to present the findings to teachers in an accessible and relevant manner. A key challenge is to design feedback that is scientifically valid, practically relevant, and ultimately contributes to the development of teaching (Brühwiler and Leutwyler, 2020). It has not yet been conclusively determined what formats and content make feedback particularly effective. Based on these research gaps, the overarching aim of the INTAKT project is to develop a comprehensive set of instruments for assessing pedagogical interaction quality. The aim of the present pilot study was to test the newly developed multi-perspective instruments with regard to their applicability and comprehensibility, and to generate initial insights into perspective differences between teachers, students, and observers, and to explore how the resulting multi-perspective data can be translated into accessible and pedagogically meaningful feedback for teachers.

The development of the INTAKT instrument follows a deliberate staged process. The present pilot study constitutes the initial stage: it examines (a) the feasibility of a multi-method, multi-perspective approach in classroom research, (b) cross-perspective differences across the 18 constructs of the INTAKT taxonomy, (c) how the resulting data can be translated into accessible feedback formats for teachers, and (d) preliminary psychometric diagnostics that inform indicator revision for the main study. The subsequent stage—full psychometric validation of the revised instrument—is the focus of the ongoing INTAKT main study ($n = 120$ classes, data collection completed autumn 2026).

For the organizational domain, systematic factor-analytic and configural-invariance analyses are reported in a parallel companion paper (Eckstein et al., under Review), which also documents the iterative indicator-selection procedure that yielded a reduced indicator set for the main study. The present paper draws on the same pilot data set but uses the pilot version of the instrument with all originally formulated indicators per construct. Findings are reported with a focus on feasibility and cross-perspective patterns; psychometric diagnostics are presented as initial pilot evidence on the full pilot instrument rather than as instrument validation.

Within this staged framework, the present pilot study applies a multi-perspective approach to capture behavior and subjective experience and explores how cross-perspective differences vary across constructs and between classrooms identified as having comparatively high or low levels of interaction quality. This approach differs from common psychometric validation procedures by focusing on the feasibility of the research design in general and the structuring of multi-perspective data in pedagogically meaningful ways. The overarching question is therefore whether such a multimethod, multi-perspective design is feasible in classroom research and yields data that are

meaningful for teaching development. It thus addresses the following research questions:

1. Can the quality of pedagogical interactions be systematically and feasibly assessed from teacher, student, and observer perspectives using the newly developed INTAKT multi-perspective instrument?
2. To what extent do teachers, students, and external observers differ in their assessments of pedagogical interaction quality, and how do these differences vary across constructs of the INTAKT taxonomy and between classrooms with higher vs. lower observer-rated interaction quality?
3. Are the data collected with the multi-perspective INTAKT instrument suitable for representation and interpretation using effect sizes and visual formats for the purpose of providing feedback to teachers?

2 Method

2.1 Design

The present pilot study is part of the Pedagogical Interaction – INTAKT research project (SNSF project number: 100019_219487), funded by the Swiss National Science Foundation. Data on pedagogical interaction quality for the present pilot study were collected using two main methods: systematic classroom observation (external perspective) and surveys filled out by teachers and students (internal perspective).

In each class, two consecutive lessons in a core subject were observed. Two lessons thus represent a pragmatic compromise for a pilot study balancing reliability and feasibility: whereas one lesson suffices for stable ratings of classroom management and climate, constructs related to cognitive activation require more observations (Praetorius et al., 2014). Accordingly, observation-based ratings can be expected to be more stable for organizational and climate-related constructs than for those in the knowledge co-construction domain. To examine ecological validity, teachers and students were asked to assess how typical the observed lessons were compared to the usual classroom events across all three teaching domains. These representativity ratings serve to contextualize the observed lessons and to support a more grounded interpretation of the observation-based findings. The observations were conducted live by two trained members of the research team, each observing independently their assigned share of classes. Both raters collaboratively developed the structured coding scheme in close alignment with the theoretical framework (see Section 1.3), and initial coding consensus was established through repeated calibration sessions using prior video material prior to data collection. To further refine the coding procedure and ensure consistent application, both raters observed one class simultaneously during data collection, established consensus coding for this class, and used these insights to inform any necessary revisions of codings in other classes. Consistent with the pilot study's focus on feasibility, each of the remaining classes was observed by only one rater; formal inter-rater

reliability assessment will be conducted in the main study through simultaneous coding of a subset of lessons.

Immediately after the observed lessons, the teachers and students completed a standardized online survey. Teachers, students and their parents were informed that participation was voluntary and that their responses would be treated confidentially, and they provided their written informed consent prior to participation in the study. The research project was reviewed and approved by the Ethics Committee of the University of Teacher Education Bern.

2.2 Instrument

The new multi-perspective instrument for assessing pedagogical interaction quality operationalizes the INTAKT taxonomy introduced above, ensuring that teachers', students', and observers' perspectives on interaction quality are captured in a comparable way.

All three perspectives assessed the same set of constructs within the three teaching domains of pedagogical interaction – *organization*, *climate*, and *knowledge co-construction* (see Table 1; Section 1.3). Each construct was operationalized through multiple items (5–11 per construct), resulting in a total of 130 indicators across 18 constructs, covering behavioral constructs, relational constructs, and general subjective experience constructs. To support construct validity and cross-perspective comparability, indicators of the observation coding scheme and surveys had equivalent meaning in perspective-adapted wording—for example, teachers rated their own classroom management (e.g., “I explained the steps clearly”), while students and observers rated the same construct from their perspective (“The teacher explained the steps clearly”). Beyond actor-specific behavioral constructs, the instrument also includes relational constructs that cannot be assigned to either actor alone (e.g., *classroom disturbances*, *teacher-student relationship*), as well as general subjective experience constructs reflecting actors' overall impressions of the classroom and their own involvement (e.g., *working atmosphere*, *well-being*, *comprehensibility*). An overview of all constructs and example items per perspective is provided in the [Supplementary Appendix \(Table A1\)](#). The items were developed specifically for the INTAKT instrument, drawing on existing measures including the Teaching Quality Feedback Instrument (UFB; Fauth and Ruth-Herbein, 2021), the INSULA instrument (Wemmer-Rogh et al., 2023), and prior survey instruments from related projects (Eckstein, 2018; Wettstein et al., 2016). The observation coding scheme builds additionally on the Behavioral Observation System for School Settings (BASYS; Wettstein, 2008). A central aim of the present pilot study was to examine whether the items are applicable and comprehensible across perspectives—not to establish full psychometric validation, which requires larger samples. In line with the staged development process outlined in Section 1.4, the present paper draws on the pilot version of the instrument with all initially compiled indicators per construct, while a complementary paper by Eckstein et al. (under review) reports the iterative indicator-selection procedure and the resulting reduced indicator set for the main study, restricted to constructs of the organizational

TABLE 2 Internal consistency (Cronbach's α) and intraclass correlations.

Construct	Teacher		Student			Observer
	α	αL_1	αL_2	ICC(1)	ICC(2)	α
Structuring	0.77	0.85	0.96	0.22	0.78	0.38
Monitoring	0.89	0.84	0.93	0.15	0.69	0.89
Academic engagement	0.95	0.89	0.99	0.26	0.82	0.98
Disciplinary behavior	0.88	0.86	0.87	0.15	0.70	0.96
Classroom disturbances	0.93	0.89	0.93	0.11	0.61	0.97
Working atmosphere	0.90	0.90	0.98	0.18	0.74	0.96
Emotional support	0.78	0.89	0.96	0.27	0.83	0.77
Teacher aggression	0.79	0.87	0.97	0.30	0.85	0.63
Prosocial behavior (students)	0.94	0.87	0.95	0.18	0.74	0.95
Student aggression	0.86	0.92	0.97	0.26	0.82	0.87
Prosocial behavior (teacher)	0.94	0.92	0.95	0.07	0.49	0.93
Teacher–student relationship	0.86	0.88	0.99	0.36	0.88	0.94
Wellbeing	0.97	0.94	0.99	0.26	0.82	0.96
Cognitive activation	0.79	0.86	0.95	0.16	0.71	0.89
Instructional support	0.71	0.90	0.99	0.28	0.83	0.83
Learning commitment	0.90	0.92	0.97	0.22	0.78	0.97
Teaching–learning processes	0.86	0.83	0.97	0.21	0.77	0.90
Intelligibility	0.93	0.90	0.98	0.23	0.79	0.95

α = Cronbach's alpha (internal consistency) for teacher, student (individual and classroom-aggregated level), and observer perspectives; ICC(1) = intraclass correlation coefficient based on the manifest scale mean; ICC(2) = reliability of the classroom mean (Spearman-Brown formula based on average classroom size).

domain. This complementary paper additionally provides systematic factor-analytic and configural-invariance analyses for that domain. Internal consistency (Cronbach's alpha) and intraclass correlation coefficients [ICC(1) and ICC(2)] are reported in Table 2; factor analyses, included as preliminary psychometric diagnostics for all 18 constructs across perspectives, are reported in Table 3. A comprehensive examination of factorial structure, internal consistency, and measurement invariance across all 18 constructs and three perspectives is reserved for the main study.

All items—across perspectives and methods—were rated on a 7-point Likert scale from “Not at all true” (1) to “Completely true” (7). The students' survey included additional personal information, such as the students' age, type of secondary school, grade level, and family background (e.g., parents' level of education). Similarly, teachers were asked to provide information about their age, employment as a teacher, attitudes toward their profession, and job satisfaction.

2.3 Sample

School principals from the cantons of Zurich, Bern, Obwalden, and Nidwalden were contacted based on cantonal school registries and asked to invite teachers to participate in the pilot study. Teachers registered voluntarily online, and participation required informed consent from teachers, students, and parents. The resulting sample is therefore a voluntary

convenience sample, with self-selection of participating teachers and classes; implications for generalizability are addressed in the Limitations section.

The sample included 16 secondary school classes and their teachers. The teachers' average age was 40.59 years ($n = 14$, $SD = 11.56$; 62.5% female and 37.5% male; no diversity information available), and they had an average professional experience of 11.04 years ($n = 14$, $SD = 10.04$). The average level of employment was 79.52% ($n = 15$, $SD = 18.38$). Missing values in teacher demographics resulted from non-response to optional survey items.

A total of $N = 211$ students participated in the survey. After data collection, the survey data were examined for response quality and outliers. Mechanical response patterns were identified based on the standard deviation within the scales. Cases with identical responses across multiple scales or a processing time below the threshold value (less than 15 min) were marked as potentially unreliable. This threshold was then compared with the requirement levels of the Swiss secondary school system—basic and extended requirements—to account for potential comprehension issues. Implausible variable values were coded as missing (NA) but were left in the sample so that they could still be used in the full information maximum likelihood (FIML) estimation. Furthermore, potential univariate outliers (threshold: $\pm 3 SD$) were identified and graphically examined, but all identified cases were ultimately judged to be unproblematic (Tabachnick and Fidell, 2014). This resulted in a final sample of $N = 209$ students ($n = 185$, $M_{age} = 14.38$, $SD = 0.94$; $n = 205$;

TABLE 3 Confirmatory factor analysis model Fit indices.

Construct	Teacher				Student (L ₁)				Student (L ₂)				Student (Two-level)				Observer	
	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR
Structuring	1.000	0.000	0.047	0.973	0.089	0.032	0.997	0.037	0.041	0.964	0.062	0.040	1.000	0.000	0.077			
Monitoring	0.993	0.049	0.068	0.903	0.162	0.058	0.919	0.232	0.059	0.921	0.086	0.067	0.939	0.190	0.109			
Academic engagement	0.897	0.277	0.067	0.960	0.133	0.033	0.880	0.320	0.040	0.980	0.054	0.038	0.930	0.324	0.011			
Disciplinary behavior	0.942	0.141	0.073	0.956	0.112	0.042	1.000	0.000	0.044	0.910	0.099	0.069	0.906	0.298	0.051			
Classroom disturbances	0.917	0.235	0.052	0.927	0.165	0.047	0.799	0.369	0.076	0.938	0.093	0.050	0.991	0.088	0.020			
Working atmosphere	0.810	0.303	0.087	0.961	0.134	0.034	0.915	0.248	0.054	0.963	0.080	0.044	0.822	0.438	0.054			
Emotional support	0.369	0.413	0.160	0.966	0.126	0.035	0.766	0.472	0.083	0.955	0.082	0.048	0.555	0.435	0.180			
Teacher aggression	0.980	0.110	0.051	0.912	0.138	0.053	0.885	0.278	0.038	0.906	0.092	0.067	1.000	0.000	0.553			
Prosocial behavior (students)	0.960	0.171	0.046	1.000	0.000	0.023	0.812	0.334	0.088	0.986	0.038	0.027	0.931	0.226	0.053			
Student aggression	0.492	0.414	0.142	0.847	0.181	0.064	0.777	0.337	0.050	0.845	0.117	0.076	0.527	0.437	0.226			
Prosocial behavior (teacher)	0.709	0.554	0.053	0.991	0.065	0.021	0.911	0.253	0.038	0.990	0.043	0.027	0.980	0.110	0.045			
Teacher-student relationship	0.838	0.244	0.092	0.981	0.092	0.026	0.929	0.274	0.026	0.962	0.073	0.044	0.842	0.359	0.056			
Wellbeing	1.000	0.000	0.031	0.943	0.205	0.030	0.847	0.424	0.051	0.942	0.120	0.041	0.872	0.343	0.040			
Cognitive activation	0.745	0.301	0.133	0.968	0.095	0.032	0.976	0.103	0.053	0.954	0.070	0.044	0.607	0.474	0.105			
Instructional support	0.707	0.313	0.105	0.968	0.122	0.028	0.970	0.185	0.018	0.979	0.057	0.036	0.808	0.231	0.128			
Learning commitment	0.905	0.196	0.073	0.993	0.060	0.019	0.722	0.480	0.089	0.977	0.064	0.025	0.974	0.157	0.021			
Teaching-learning processes	0.690	0.392	0.140	0.967	0.095	0.038	0.924	0.228	0.066	0.960	0.061	0.052	0.778	0.395	0.131			
Intelligibility	0.879	0.233	0.068	1.000	0.000	0.015	0.980	0.115	0.029	1.000	0.000	0.023	0.926	0.217	0.043			

CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual. For the two-level model, SRMR reflects within-level fit only (SRMR_{within}); between-level SRMR is not reported due to instability at the present classroom sample size (*n* = 16). Robust (MLR-scaled) values are reported where available; otherwise standard ML values are shown.

48.8% female, 47.37%, male, 1.91% diverse) from different Swiss secondary school tracks ($n = 201$, 24.88% basic requirements [GA], 71.29% extended requirements [eA]) and different grade levels ($n = 207$, 47.85% seventh grade, 34.93% eighth grade, 16.27% ninth grade). Missing values in student demographics resulted from non-response to optional items. Item-level missing data in the interaction quality scales were minimal and handled using FIML estimation in all analyses.

2.4 Analytic strategy

The analytic strategy for this pilot study reflects its feasibility-focused approach. Given the small classroom sample ($n = 16$), inferential testing at the classroom level would be statistically underpowered and risk over-interpretation of pilot data. We therefore based our analyses on descriptive measures, absolute differences (Δ), and standardized effect sizes (adapted Cohen's d , hereafter d') rather than null-hypothesis significance testing, in line with recommendations for evaluation research with small samples (Lind, 2012; Valentine et al., 2015). Following the staged development process described in Section 1.4, the analytic strategy comprises three main approaches: (1) descriptive analyses and perspective differences, (2) standardization and effect sizes, and (3) contrast group comparison for feedback illustration. In addition to these feasibility-focused analyses, initial psychometric diagnostics were conducted in parallel to inform instrument revision for the main study. This analytical sequence—with substantive analyses preceding formal psychometric evaluation—was necessary because we had committed to providing participating teachers with timely feedback reports on their classroom interactions, which required prioritizing the completion of descriptive and comparative analyses.

All analyses were performed in R (R Core Team, 2026) using the lavaan (Rosseel, 2012), psych (Revelle, 2025) and lme4 (Bates et al., 2015) packages for data analysis, with tidyverse (Wickham et al., 2019) for data preparation.

2.4.1 Descriptive analyses and perspective differences

The initial analytical strategy involved descriptive analyses of the 18 constructs of pedagogical interaction quality at both the sample and classroom level, as well as the computation of absolute perspective differences (Δ) and corresponding effect sizes (adapted Cohen's d , hereafter d') between teachers, students, and observers at the classroom level. These classroom-based analyses were the basis for the teacher feedback reports.

For functional constructs, values of 1–2 are considered low, 3–5 moderate, and 6–7 high; for dysfunctional constructs (classroom disturbances, teacher aggression, student aggression), values of 1–2 indicate high interaction quality, 3–5 moderate, and 6–7 low interaction quality.

Δ refers to the unsigned mean difference between two perspectives (e.g., $|M_T - M_S|$). Thresholds for interpreting d' were set at <0.30 (small), 0.30 – 0.59 (medium, marked with *), and ≥ 0.60 (large, marked with **), following Hattie (2012). The asterisks refer to d' only; Δ is reported without asterisks.

2.4.2 Standardization and effect sizes

To account for the data structure—with teacher and observer perspectives providing only single class-level values and therefore lacking individual variance estimates—pairwise pooled standard deviations were applied. Furthermore, the absolute perspective differences were visualized to systematically represent perceptual discrepancies, with d' contextualizing the relative magnitude of discrepancies across constructs and Δ capturing their direct extent at the classroom level. Additionally, standardized values (z -values) were computed to assess the intra-class perspective differences in relation to the total sample.

2.4.3 Contrast group comparison for feedback illustration

In a subsequent step, an anonymized sample results report was generated, which summarized the pedagogical interaction quality of two contrast groups. These contrast groups were created based on observer ratings of the 18 constructs specified by the INTAKT taxonomy across the teaching domains of organization, climate, and knowledge co-construction (see Table 1; see also Section 1.3). For each class, the sum of the 15 functional constructs was calculated, and the sumscore of the three dysfunctional constructs was subtracted. This composite score served to determine the overall level of interaction quality per class. The three classes with the highest and the three classes with the lowest summary values were then combined into two contrast groups and labeled as the *Top-3 High-Interaction Contrast Group (Top-3)* and the *Bottom-3 Low-Interaction Contrast Group (Bottom-3)*, respectively. In line with the pilot study's focus on examining the feasibility of multi-perspective feedback, these contrast groups serve to illustrate typical patterns and the range of interaction quality within the sample; they do not support causal inferences or generalizations beyond this sample. For further descriptive purposes and classification while preserving anonymity, the mean values of the two contrast groups were also z -standardized, that is, set in relation to the mean value of the total sample ($M = 0$) per perspective. The resulting z -values enable an evaluation of the extent to which individual features in the sample classes deviate from the mean.

2.4.4 Initial psychometric diagnostics

Initial psychometric diagnostics were conducted in parallel to the substantive analyses to explore the pilot version of the instrument and inform the revision of indicators for the main study. Internal consistency was estimated using Cronbach's α at four levels: teacher ratings ($n = 16$); student ratings at the individual level (L_1 ; $n = 209$); student ratings aggregated to the classroom level (L_2 ; $n = 16$); and observer ratings ($n = 16$). To evaluate the suitability of multilevel analysis for student data, the intraclass correlation coefficients were computed: ICC(1) was derived from a two-level null model, and ICC(2) was computed using the Spearman–Brown formula based on the average classroom size. Confirmatory factor analyses (CFAs) were conducted separately for each of the 18 constructs, with five model specifications examined for each—single-level CFAs based on teacher, student (L_1 and L_2), and observer ratings; and

a two-level CFA based on student data, with within-level SRMR ($SRMR_w$) reported as the primary fit index for this model; between-level SRMR ($SRMR_b$) is not reported, as estimates are structurally unstable at the present classroom sample size ($n = 16$). The models were estimated using maximum likelihood with robust standard errors (MLR) and full information maximum likelihood (FIML) to account for missing data. In the few two-level models with difficulty in estimation, between-level residual variances were fixed to zero. The model fit was evaluated against established thresholds: $CFI \geq .90$, $RMSEA \leq .08$, and $SRMR \leq .08$ (Hu and Bentler, 1999). Robust (MLR-scaled) fit indices are reported where available; otherwise, standard ML values are shown. Single-level CFAs based on teacher and observer data ($n = 16$) are reported for purposes of completeness but should be interpreted cautiously due to the small sample size at the classroom level.

3 Results

3.1 Initial psychometric properties of the pilot instrument

This section first reports the initial psychometric properties of the pilot instrument across the 18 constructs and three perspectives. In line with the staged development process described in Section 1.4, these analyses are reported as pilot diagnostics on the full pilot instrument rather than as instrument validation. Their purpose is both to characterize the constructs as used in the present pilot study and to inform indicator revision for the main study. Detailed values are reported in Tables 2 (internal consistency and intraclass correlations) and 3 (confirmatory factor analyses).

Internal consistency was acceptable to excellent across the methodologically most informative models. At the student individual level (L_1 ; $n = 209$), α ranged from .83 to .94 across all 18 constructs, indicating consistent reliability of student ratings. At the classroom-aggregated student level (L_2 ; $n = 16$), α values were even higher (.87 to .99), reflecting high levels of agreement among students within classes when accounting for individual variance. For teacher ratings ($n = 16$), alpha ranged from .71 to .97, with the lowest values recorded for *instructional support* (.71) and *cognitive activation* (.79). Observer ratings revealed the broadest range (.38 to .98), with most values exceeding .77; notable exceptions were *structuring* ($\alpha = .38$) and, to a lesser extent, *teacher aggression* ($\alpha = .63$), where the indicator wording has been considered for the main study.

Intraclass correlation coefficients on student data indicated substantial classroom-level variance and supported the appropriateness of multilevel analyses. ICC(1) values ranged from .07 (*prosocial behavior towards the teacher*) to .36 (*teacher-student relationship*), with predominant values falling between .15 and .30. ICC(2) values, reflecting the reliability of the classroom mean, ranged from .49 to .88; the majority of constructs exceeded .70, indicating that classroom-aggregated student ratings reliably varied between classes. Consequently, these patterns support the use of classroom-level aggregates of student ratings in the subsequent cross-perspective analyses.

Confirmatory factor analyses on student data generally supported the unidimensional structure expected per construct and perspective on the most methodologically informative models.

For the two-level CFA on student data, 17 of 18 constructs met the CFI criterion ($\geq .90$); however, RMSEA values were partly elevated (with 10 of 18 below .08), reflecting limits of model estimation given the modest classroom sample. $SRMR_w$ values were below .08 across all 18 constructs in the two-level model.

The single-level CFA on student ratings revealed a comparable pattern. On the student-level models, weaker fit was observed specifically for *student aggression* ($CFI = .85$) and *wellbeing* (with elevated RMSEA values); these constructs require closer examination during indicator revision for the main study.

Single-level CFAs on teacher and observer data were estimated for purposes of completeness. However, as stated in the method section (see Section 2.4.1), these findings must be interpreted in light of the very small sample size at the classroom level ($n = 16$). Consequently, the fit indices demonstrated instability, with certain constructs attaining adequate fit (e.g., *wellbeing* on teacher data, $CFI = 1.00$; *teacher aggression* on teacher data, $CFI = .98$). Conversely, several models yielded fit values considerably below conventional thresholds, particularly for *emotional support*, *student aggression*, and *cognitive activation*. These results reflect the structural limits of CFA estimation at $n = 16$ rather than fundamental problems of the constructs themselves; the corresponding student-level and two-level models indicate that the underlying constructs may be estimable when sufficient information is available.

In sum, the psychometric diagnostics support the use of the pilot instrument for the descriptive and cross-perspective analyses reported below; while emphasizing specific constructs whose indicators require targeted revision for the main study.

3.2 Descriptive findings at the sample level

The following results section begins with a descriptive overview of the perceived quality of pedagogical interaction at the sample level. This serves as a foundation for the subsequent classroom-level analyses of cross-perspective differences. For descriptive purposes, we apply the following convention when referring to ratings as low, moderate, or high. For functional constructs, values of 1–2 are considered *low*, 3–5 *moderate*, and 6–7 *high*; for dysfunctional constructs, higher values indicate lower interaction quality, whereas lower values indicate higher interaction quality. An overview of the mean values and standard deviations for all 18 constructs across the three perspectives is provided in Table 4.

In the domain of classroom *organization*, *structuring* is rated consistently high across all three perspectives: teachers ($M = 6.03$, $SD = 0.53$), students ($M = 6.11$, $SD = 0.78$), and observers ($M = 6.39$, $SD = 0.50$). *Monitoring* is rated slightly higher by observers ($M = 5.58$, $SD = 1.04$) and students ($M = 5.57$, $SD = 0.94$) compared to teachers ($M = 5.01$, $SD = 0.63$). In line with this, ratings for *classroom disturbances* are generally low, particularly from the observer perspective ($M = 2.75$, $SD = 1.76$), whereas teachers ($M = 3.59$, $SD = 1.26$) and students ($M = 3.46$, $SD = 1.35$) report more pronounced disturbances. Teachers assess students' *academic engagement* as

TABLE 4 Descriptive statistics at the sample level.

Construct	Teacher		Student		Observer	
	M	SD	M	SD	M	SD
Structuring	6.03	0.53	6.11	0.78	6.39	0.50
Monitoring	5.01	0.63	5.57	0.94	5.58	1.04
Academic engagement	4.73	0.87	5.18	0.95	5.21	1.78
Disciplinary behavior	4.63	0.96	4.98	0.96	4.48	0.61
Classroom disturbances	3.59	1.26	3.46	1.35	2.75	1.76
Working atmosphere	4.94	0.67	5.70	1.05	5.50	1.58
Emotional support	5.34	0.49	5.79	1.11	6.00	0.87
Teacher aggression	1.40	0.53	1.76	0.93	1.25	0.52
Prosocial behavior (students)	4.91	0.98	5.73	0.86	6.00	1.30
Student aggression	1.64	0.66	2.32	1.24	1.97	1.38
Prosocial behavior (teacher)	5.49	0.72	6.06	0.79	5.90	1.10
Teacher–student relationship	5.16	0.49	5.47	1.14	5.94	1.19
Wellbeing	5.31	0.74	5.56	1.41	5.43	1.50
Cognitive activation	3.81	1.14	5.38	1.11	5.10	1.31
Instructional support	4.84	0.61	5.63	1.17	5.83	0.92
Learning commitment	4.58	0.77	5.04	1.15	5.20	1.59
Teaching–learning processes	4.72	0.75	5.21	0.99	5.16	1.37
Intelligibility	5.04	0.69	5.55	1.03	5.38	1.48

M, mean; SD, standard deviation. Constructs are sorted by teaching domain within the INTAKT taxonomy.

moderate ($M = 4.73$, $SD = 0.87$), while students ($M = 5.18$, $SD = 0.95$) and observers ($M = 5.21$, $SD = 1.78$) perceive it as somewhat higher. In contrast, *disciplinary behavior* is rated as moderate across all three perspectives ($M_T = 4.63$, $SD = 0.96$; $M_S = 4.98$, $SD = 0.96$; $M_{Obs} = 4.48$, $SD = 0.61$). Overall, the data suggest strong convergence between perspectives on *structuring* and *monitoring*, whereas *classroom disturbances* and *academic engagement* show more variability. Ratings for *disciplinary behavior* are broadly aligned and consistently moderate.

In the domain of *climate*, observers ($M = 6.00$, $SD = 0.87$) and students ($M = 5.79$, $SD = 1.11$) perceive teachers' *emotional support* as comparatively high. Teachers report slightly lower values ($M = 5.34$, $SD = 0.49$), but still reflect a clearly favorable perception. A similar pattern appears for the *teacher-student relationship*, with observers ($M = 5.94$, $SD = 1.19$), students ($M = 5.47$, $SD = 1.14$), and teachers ($M = 5.16$, $SD = 0.49$) perceiving it as moderate to relatively high. Students' *prosocial behavior* is rated relatively high across all perspectives, both among peers ($M_{Obs} = 6.00$, $SD = 1.30$; $M_S = 5.73$, $SD = 0.86$; $M_T = 4.91$, $SD = 0.98$) and toward teachers ($M_{Obs} = 5.90$, $SD = 1.10$; $M_S = 6.06$, $SD = 0.79$; $M_T = 5.49$, $SD = 0.72$). *Aggressive student behavior* receives low ratings from all three perspectives ($M_T = 1.64$, $SD = 0.66$; $M_S = 2.32$, $SD = 1.24$; $M_{Obs} = 1.97$, $SD = 1.38$). *Teacher aggression* is also rarely reported ($M_T = 1.40$, $SD = 0.53$; $M_S = 1.76$, $SD = 0.93$; $M_{Obs} = 1.25$, $SD = 0.52$).

In the domain of *knowledge co-construction*, students ($M = 5.38$, $SD = 1.11$) and observers ($M = 5.10$, $SD = 1.31$) rate *cognitive activation* higher than teachers ($M = 3.81$, $SD = 1.14$). A similar pattern emerges for *instructional support*, with observers ($M = 5.83$, $SD = 0.92$) and students ($M = 5.63$,

$SD = 1.17$) again assigning higher ratings than teachers ($M = 4.84$, $SD = 0.61$). *Teaching-learning processes* are rated as moderate to rather high across all perspectives ($M_{Obs} = 5.16$, $SD = 1.37$; $M_S = 5.21$, $SD = 0.99$; $M_T = 4.72$, $SD = 0.75$). Perceived *comprehensibility* follows a similar pattern, with moderate to high ratings from observers ($M = 5.38$, $SD = 1.48$), students ($M = 5.55$, $SD = 1.03$), and teachers ($M = 5.04$, $SD = 0.69$).

These sample-level findings provide an initial descriptive overview of how pedagogical interaction quality is perceived across the three perspectives. In the following sections, we turn to classroom-level analyses, focusing on perspective differences and their implications for feedback.

3.3 Cross-perspective perceptions at the classroom level

The following classroom-level analyses focus on perspective differences in two contrast groups formed based on observer ratings. As described in Section 2.4.3, the *High-Interaction Contrast Group (Top-3)* comprises the three classes with the highest observer-rated interaction quality, while the *Low-Interaction Contrast Group (Bottom-3)* comprises the three classes with the lowest observer-rated interaction quality. These contrast groups serve to illustrate the range of interaction quality patterns within the sample and to explore whether perspective differences vary systematically across classroom contexts.

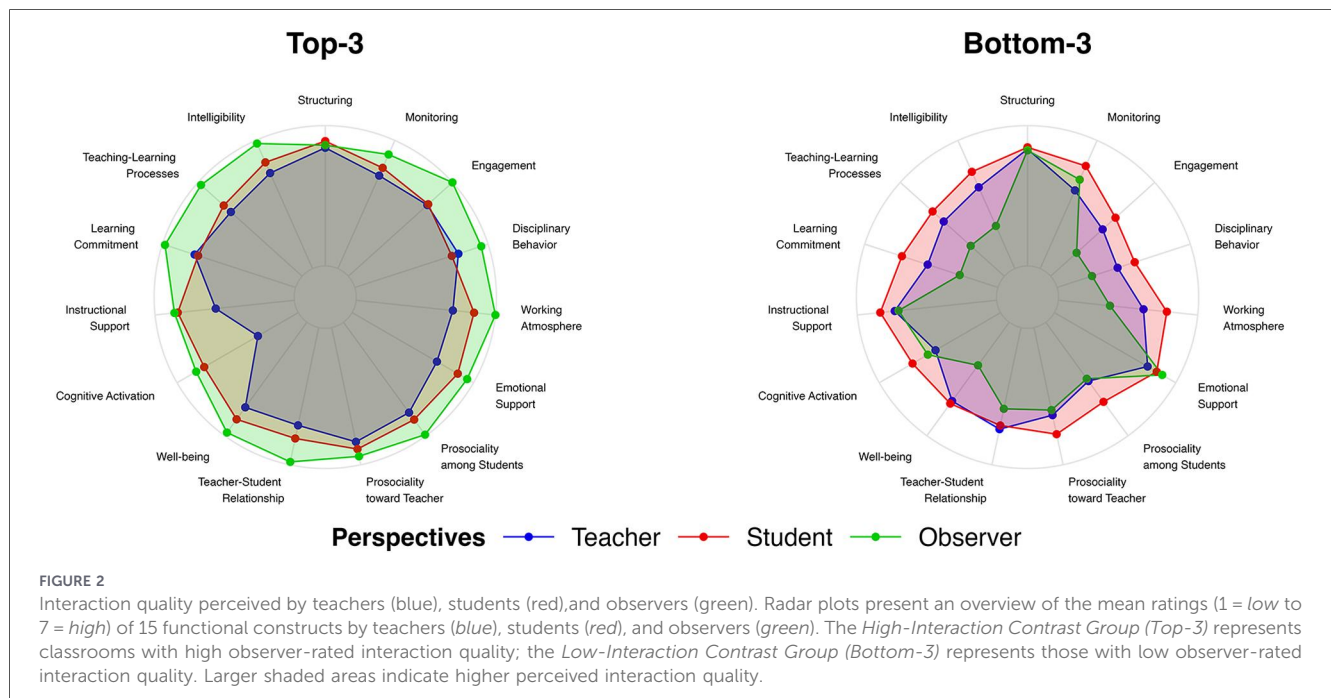


Figure 2 provides an initial overview of the assessment patterns of teachers (blue), students (red), and observers (green) regarding pedagogical interaction quality in the two contrast groups and serves as the basis for the subsequent interpretation of perspective differences. It covers the 15 functional constructs of the INTAKT taxonomy arranged along individual axes of a radar plot. The value range extends from 1.0 (low interaction quality, inner circle) to 7.0 (high interaction quality, outer circle). Assessments from each perspective are marked as points on the respective axes and connected to form a shaded area. These areas illustrate the overall picture of pedagogical interaction quality from each perspective: the larger the area, the higher the assessed quality. In the *High-Interaction Contrast Group (Top-3)*, the areas are clearly pronounced in almost all constructs and from all perspectives, indicating consistently high perceived pedagogical interaction quality. Only specific constructs (e.g., *cognitive activation* from the teacher perspective or *disciplinary behavior* from the student and observer perspectives) show noticeable dips, which will be addressed below (see Section 3.3). In the *Low-Interaction Contrast Group (Bottom-3)* the areas appear overall smaller, yet values are not consistently low across all constructs. From all three perspectives, the interaction quality tends to lie in the middle range or slightly below. Particularly low values appear in the ratings of students' *disciplinary behavior* and *academic engagement*. These variations suggest that the *Low-Interaction Contrast Group (Bottom-3)* does not reflect a uniformly negative picture but rather reveals a mixed or differentiated profile of pedagogical interaction quality.

3.4 Cross-perspective patterns by teaching domain

The following analyses examine classroom-level perspective differences separately for each of the three teaching domains. Figures 3–6 present mean absolute ratings for pedagogical

interaction quality constructs, rated by teachers, students, and observers on a 1.0 (low interaction quality) to 7.0 (high interaction quality) scale. Effect sizes for perspective differences are indicated by * (medium differences, $0.3 \leq d' < 0.6$) or ** (large differences, $d' \geq 0.6$).

3.4.1 Cross-perspective patterns in classroom organization

Organizational constructs showed three salient patterns (see Figures 3 and 4). First, *structuring* received consistently high ratings across all three perspectives in both contrast groups, with minimal discrepancies. Second, *monitoring* revealed consistent teacher underestimation relative to students and observers across both groups, with this discrepancy considerably larger in the *Low-Interaction Contrast Group (Bottom-3)* ($\Delta_{T-S} = 1.14$, $d' = 0.39^*$; $\Delta_{T-O} = 1.89$, $d' = 0.79^{**}$) than in the *High-Interaction Contrast Group (Top-3)* ($\Delta_{T-S} = 0.50$).

Third, *classroom disturbances* and *working atmosphere* showed contrasting cross-perspective patterns between groups. For *classroom disturbances*, teacher-student alignment was moderate in the *High-Interaction Contrast Group (Top-3)* ($M_T = 2.78$, $M_S = 3.40$) but higher in the *Low-Interaction Contrast Group (Bottom-3)* ($M_T = 4.72$, $M_S = 4.35$), with both perspectives reporting elevated levels in the latter group. However, teacher-observer discrepancies followed opposite patterns: in the *High-Interaction Contrast Group (Top-3)*, teachers reported higher disturbances than observers ($\Delta_{T-O} = 1.72$, $d' = 0.47^*$), while in the *Low-Interaction Contrast Group (Bottom-3)*, teachers reported lower disturbances than observers ($\Delta_{T-O} = 0.72$, $d' = 1.12^{**}$). For *working atmosphere*, large teacher-observer discrepancies emerged in both groups, though in opposite directions: in the *High-Interaction Contrast Group (Top-3)*, observers rated it as maximally high while teachers rated it more moderately ($M_{Obs} = 7.00$ vs. $M_T = 5.17$, $\Delta_{T-O} = 1.83$, $d' = 1.19^{**}$);

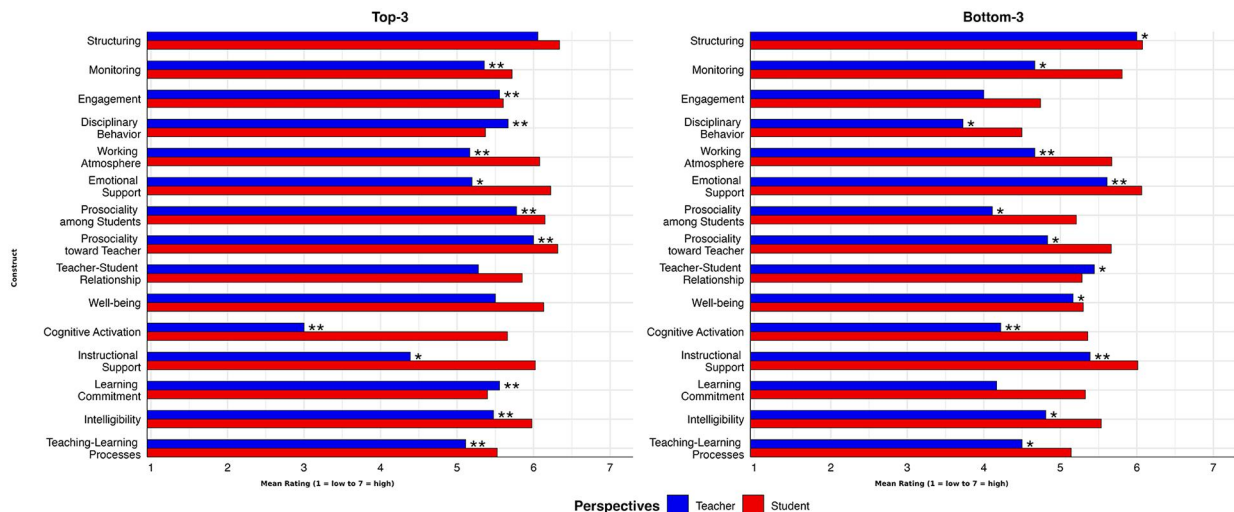


FIGURE 3

Absolute ratings by teachers (blue) and students (red). Bars present mean absolute ratings (1 = low to 7 = high) from teachers (blue) and students (red) for 15 functional constructs of pedagogical interaction quality. The *High-Interaction Contrast Group (Top-3)* represents classrooms with high observer-rated interaction quality, and the *Low-Interaction Contrast Group (Bottom-3)* represents those with low observer-rated interaction quality. Asterisks indicate standardized effect size categories for perspective differences: * = medium ($0.30 \leq d' < 0.60$), ** = large ($d' \geq 0.60$).

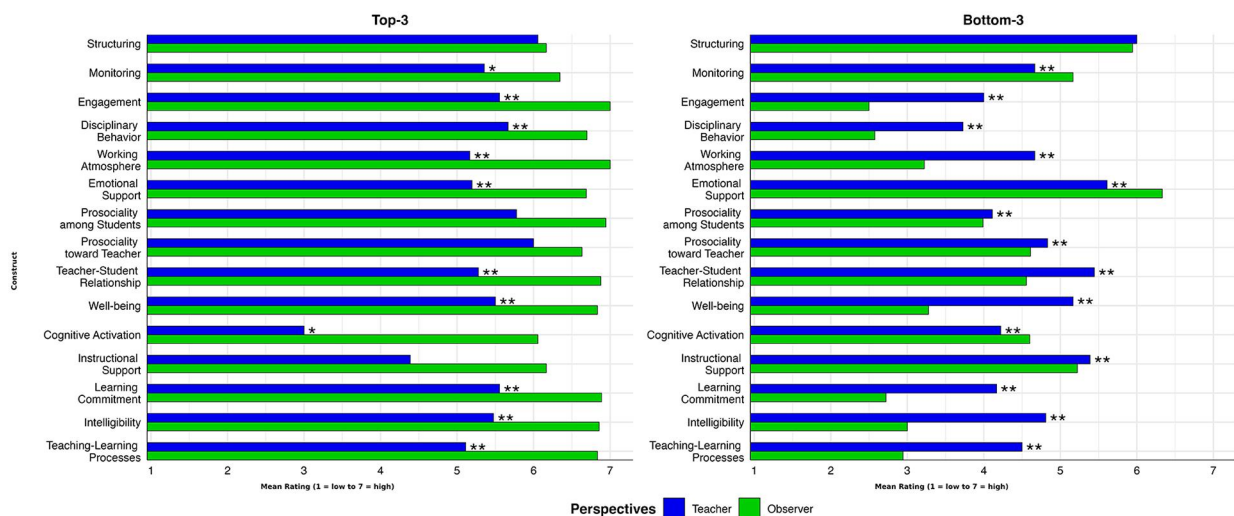


FIGURE 4

Absolute ratings by teachers (blue) and observers (green). Bars present mean absolute ratings (1 = low to 7 = high) from teachers (blue) and observers (green) for 15 functional constructs of pedagogical interaction quality. The *High-Interaction Contrast Group (Top-3)* represents classrooms with high observer-rated interaction quality, and the *Low-Interaction Contrast Group (Bottom-3)* represents those with low observer-rated interaction quality. Asterisks indicate standardized effect size categories for perspective differences: * = medium ($0.30 \leq d' < 0.60$), ** = large ($d' \geq 0.60$).

in the *Low-Interaction Contrast Group (Bottom-3)*, teachers rated it higher than observers ($M_T = 4.67$ vs. $M_{Obs} = 3.00$, $\Delta_{T-O} = 1.67$, $d' = 1.17^{**}$).

Across the organizational domain, perspective alignment varied with the inferential demands of the construct. *Structuring* revealed strong convergence across perspectives, while *monitoring*, *classroom disturbances* and *working atmosphere* revealed larger differences that sometimes reversed in direction between contrast groups.

3.4.2 Cross-perspective patterns in classroom climate

Climate-related constructs showed varying degrees of perspective alignment (see Figures 3 and 4 for functional constructs; Figures 5 and 6 for dysfunctional constructs). *Emotional support* and *teacher-student relationship* were rated moderate to high across perspectives in both groups, with relatively small teacher-student discrepancies. Teacher-observer

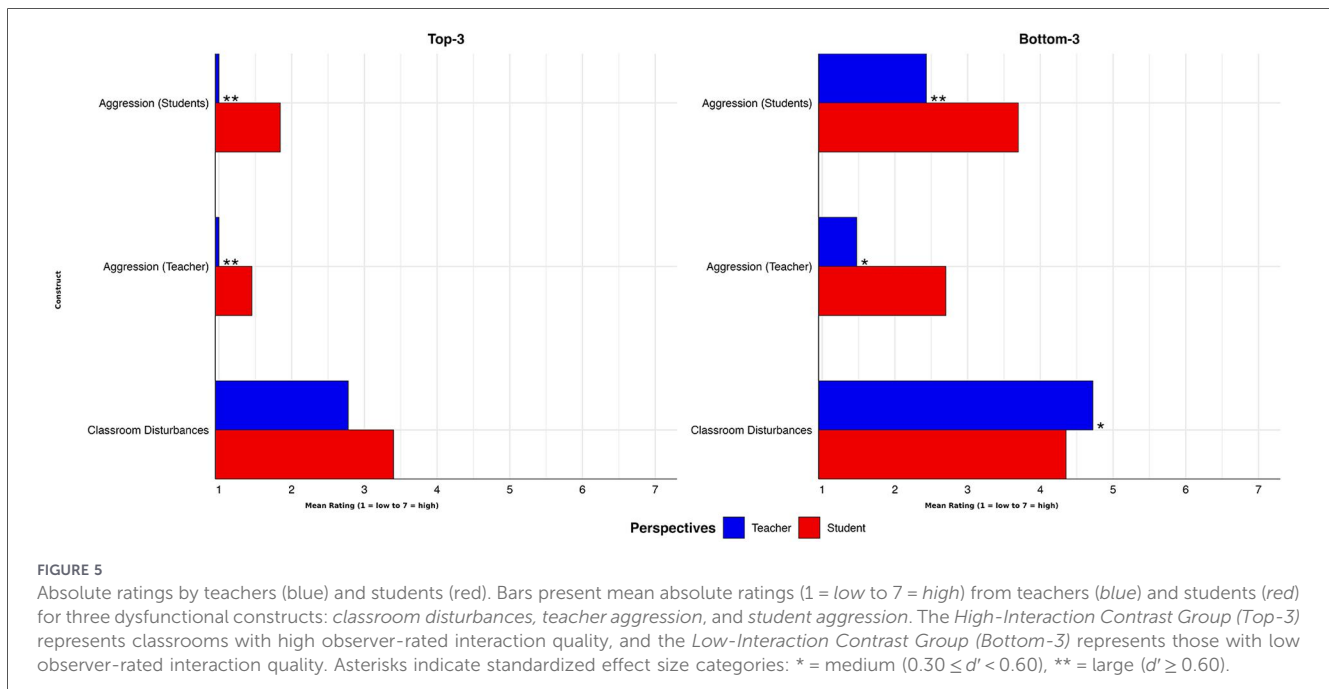


FIGURE 5

Absolute ratings by teachers (blue) and students (red). Bars present mean absolute ratings (1 = low to 7 = high) from teachers (blue) and students (red) for three dysfunctional constructs: *classroom disturbances*, *teacher aggression*, and *student aggression*. The *High-Interaction Contrast Group (Top-3)* represents classrooms with high observer-rated interaction quality, and the *Low-Interaction Contrast Group (Bottom-3)* represents those with low observer-rated interaction quality. Asterisks indicate standardized effect size categories: * = medium (0.30 ≤ d' < 0.60), ** = large (d' ≥ 0.60).

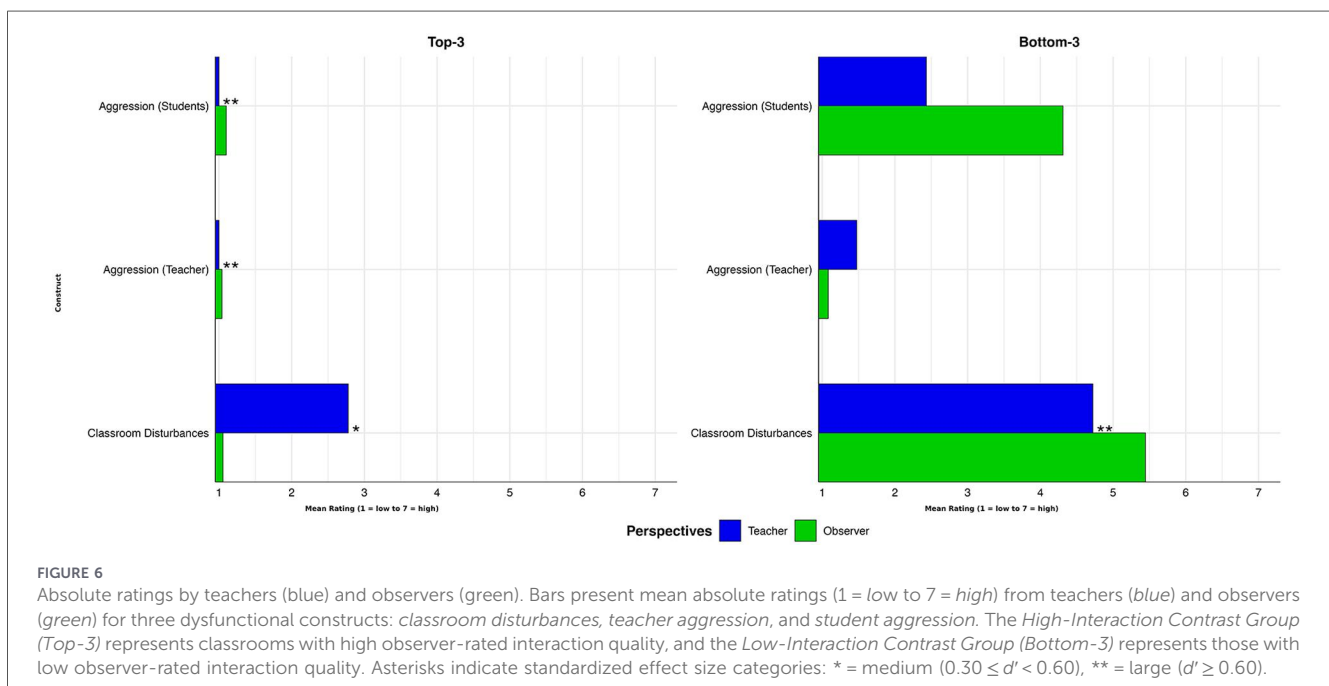


FIGURE 6

Absolute ratings by teachers (blue) and observers (green). Bars present mean absolute ratings (1 = low to 7 = high) from teachers (blue) and observers (green) for three dysfunctional constructs: *classroom disturbances*, *teacher aggression*, and *student aggression*. The *High-Interaction Contrast Group (Top-3)* represents classrooms with high observer-rated interaction quality, and the *Low-Interaction Contrast Group (Bottom-3)* represents those with low observer-rated interaction quality. Asterisks indicate standardized effect size categories: * = medium (0.30 ≤ d' < 0.60), ** = large (d' ≥ 0.60).

gaps were larger, particularly for *teacher-student relationship* in the *High-Interaction Contrast Group (Top-3)* ($\Delta_{T-O} = 1.60$, $d' = 0.98^{**}$). *Prosocial behavior* among students showed high cross-perspective alignment in the *High-Interaction Contrast Group (Top-3)*, with all three perspectives converging on elevated ratings. In the *Low-Interaction Contrast Group (Bottom-3)*, ratings were more moderate but remained relatively aligned across perspectives.

Wellbeing showed notable group-specific patterns. In the *High-Interaction Contrast Group (Top-3)*, all three perspectives

converged on high ratings ($M_T = 5.67$, $M_S = 6.00$, $M_{Obs} = 6.17$). In the *Low-Interaction Contrast Group (Bottom-3)*, however, substantial discrepancies emerged: teachers rated *wellbeing* considerably higher ($M_T = 5.17$) than both students ($M_S = 4.22$, $\Delta_{T-S} = 0.94$, $d' = 0.74^{**}$) and observers ($M_{Obs} = 3.28$, $\Delta_{T-O} = 1.89$, $d' = 1.13^{**}$).

Dysfunctional climate constructs revealed group-specific patterns (see Figures 5 and 6). In the *High-Interaction Contrast Group (Top-3)*, *aggressive behavior*—both among students and by teachers—was rated low across all perspectives. In the *Low-*

Interaction Contrast Group (Bottom-3), all three perspectives reported elevated levels, with notable discrepancies. For *student aggression*, teachers reported lower levels ($M_T = 2.43$) than both students ($M_S = 3.70$, $\Delta_{T-S} = 1.26$, $d' = 0.70^{**}$) and observers ($M_{Obs} = 4.31$, $\Delta_{T-O} = 1.88$). For *teacher aggression*, students reported higher levels ($M_S = 2.70$) than teachers themselves ($M_T = 1.48$, $\Delta_{T-S} = 1.23$, $d' = 0.50^*$), while observers' ratings fell between the two ($M_{Obs} = 1.89$).

Across the climate domain, perspective alignment varied between contrast groups. In the *High-Interaction Contrast Group (Top-3)*, perspectives largely converged; in the *Low-Interaction Contrast Group (Bottom-3)*, substantial differences emerged for *wellbeing* and *aggression*, with teachers reporting more favorable conditions than students or observers.

3.4.3 Cross-perspective patterns in knowledge co-construction

Knowledge co-construction constructs revealed the largest and most systematic perspective differences (see Figures 3 and 4). *Cognitive activation* showed pronounced teacher underestimation across both contrast groups. In the *High-Interaction Contrast Group (Top-3)*, teachers rated their cognitive activation substantially lower ($M_T = 3.00$) than both students ($M_S = 5.66$, $\Delta_{T-S} = 2.66$, $d' = 1.02^{**}$) and observers ($M_{Obs} = 6.06$, $\Delta_{T-O} = 3.06$, $d' = 0.31^*$). This pattern persisted in the *Low-Interaction Contrast Group (Bottom-3)*, with teachers again rating themselves lower than students ($\Delta_{T-S} = 0.89$, $d' = 0.42^*$) and observers. Instructional support followed a similar pattern in both groups, with teachers consistently rating their support lower than students and observers (*High-Interaction Contrast Group (Top-3)*: $\Delta_{T-S} = 1.63$, $d' = 0.54^*$; $\Delta_{T-O} = 1.44$, $d' = 0.68^{**}$).

Teaching-learning processes and *intelligibility* showed large teacher-observer discrepancies that varied by contrast group. For *teaching-learning processes*, the direction of the discrepancy reversed between groups: in the *High-Interaction Contrast Group (Top-3)*, observers rated them considerably higher than teachers ($M_{Obs} = 6.83$ vs. $M_T = 5.11$, $\Delta_{T-O} = 1.72$, $d' = 1.41^{**}$), while in the *Low-Interaction Contrast Group (Bottom-3)*, teachers rated them higher than observers ($M_T = 4.33$ vs. $M_{Obs} = 3.39$, $\Delta_{T-O} = 0.94$, $d' = 0.65^{**}$). Similarly, *intelligibility* showed minimal teacher-observer discrepancies in the *High-Interaction Contrast Group (Top-3)* but large gaps in the *Low-Interaction Contrast Group (Bottom-3)*, with teachers rating it considerably higher than observers ($M_T = 4.81$ vs. $M_{Obs} = 3.00$, $\Delta_{T-O} = 1.81$, $d' = 1.19^{**}$).

Across the knowledge co-construction domain, perspectives diverged systematically. *Cognitive activation* and *instructional support* showed persistent teacher self-underestimation across both contrast groups, while *teaching-learning processes* and *intelligibility* showed larger discrepancies primarily in the *Low-Interaction Contrast Group (Bottom-3)*.

3.5 Perspective comparison between the contrast groups and the total sample

The z-standardized values of the 15 functional core constructs, presented in Figure 7, allow for a comparative classification of the two contrast groups in relation to the total sample, thereby highlighting above- and below-average patterns of pedagogical interaction quality. In the *High-Interaction Contrast Group (Top-3)*, numerous features are rated above the sample average, reflected as in positive z-values. From the teachers' perspective, for example, the students' *commitment to learning* ($z = 1.26$) and their *disciplinary behavior* ($z = 1.08$) are clearly above the

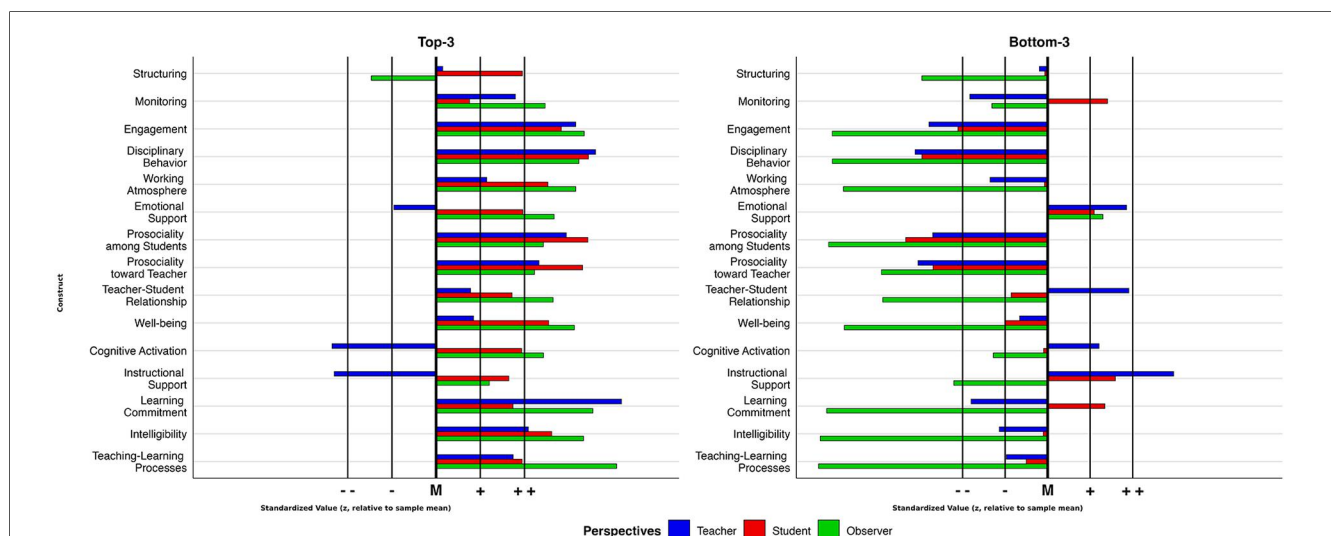


FIGURE 7

Comparison of the contrast groups with the total sample using Z values. Bars present z-standardized mean values for each construct relative to the total sample mean ($M = 0$). Values above zero indicate above-average ratings, and values below zero indicate below-average ratings. The *High-Interaction Contrast Group (Top-3)* represents classrooms with high observer-rated interaction quality, and the *Low-Interaction Contrast Group (Bottom-3)* represents those with low observer-rated interaction quality. Colors represent rating perspectives: teachers = blue, students = red, observers = green.

average for the total sample. From the students' perspective, the most notable deviations from the overall sample are evident in the *disciplinary behavior* ($z = 1.03$) and mutual *prosocial behavior of the students* ($z = 1.03$). From the observational perspective, *learning commitment* of the students shows a distinctly positive z -value of 1.06, while *teaching and learning processes* are rated even higher with $z = 1.23$. Given that the *High-Interaction Contrast Group (Top-3)* was formed based on the observation data, it is noteworthy that both teachers' and students' perspectives are consistently above the sample mean. The consistent positive assessments across all three perspectives indicate an above-average quality of pedagogical interaction in the *High-Interaction Contrast Group (Top-3)*.

Conversely, within the *Low-Interaction Contrast Group (Bottom-3)*, several features are rated below the sample average, reflected as negative z -values. Observers rated the *teaching-learning processes* ($z = -1.62$), the students' *commitment to learning* ($z = -1.56$), and the students' *prosocial behavior* toward each other ($z = -1.55$) considerably lower than in the total sample. From the teachers' perspective, the most pronounced negative deviations concern students' *disciplinary behavior* ($z = -0.93$) and their *prosocial behavior* toward the teacher ($z = -0.91$). Compared to the overall sample, students in the *Low-Interaction Contrast Group (Bottom-3)* also report lower levels of *disciplinary behavior* ($z = -0.89$) and mutual *prosocial behavior* ($z = -1.00$).

3.6 Summary of key patterns

Three overarching cross-perspective patterns emerged across the teaching domains. First, teachers consistently rated their own instructional efforts—particularly *cognitive activation* and *instructional support*—considerably lower than students and observers across both contrast groups, representing the largest and most systematic discrepancies observed. This pattern held regardless of overall classroom quality. Second, the magnitude and direction of perspective discrepancies varied systematically by contrast group for several constructs. *Monitoring* gaps widened in the *Low-Interaction Contrast Group (Bottom-3)*; *wellbeing* showed large teacher-student and teacher-observer discrepancies only in the *Low-Interaction Contrast Group (Bottom-3)*; and *teaching-learning processes* and *intelligibility* showed direction reversals between groups. Third, despite contrast groups being formed based solely on observer ratings, teacher and student ratings aligned with the observer-based classification across multiple constructs, suggesting genuine cross-perspective convergence on overall interaction quality.

4 Discussion

The overarching aim of the INTAKT research project is to assess the quality of pedagogical interactions in the classroom from multiple perspectives, considering both observable behavior and subjective experiences. Within this framework, the present pilot study explored three main questions: first, whether and how the quality of pedagogical interactions can be feasibly captured from multiple perspectives using the newly developed

instrument; second, what initial patterns emerge in how teachers, students, and external observers perceive pedagogical interaction quality in Swiss secondary school classrooms; and third, to explore how the resulting findings can be translated into feedback that is both comprehensive and useful for teachers.

First, we examined whether—and how—pedagogical interactions can be meaningfully captured from multiple perspectives. The pilot study provides preliminary evidence that this approach is feasible: the instrument was designed so that teachers, students, and observers assessed the same set of constructs using perspective-adapted items, which allowed for systematic comparisons across perspectives. However, several methodological challenges emerged. Since both teacher and observer perspectives consist of single values per class, there is no within-class variance available for the teacher and observer perspectives. To address this, we relied on pairwise pooled standard deviations to calculate effect sizes and illustrate differences in perception. Additionally, the observation data were collected by individual raters without systematic inter-rater reliability assessment, which represents a limitation that will be addressed in the main study through multiple coding procedures.

Second, we examined what patterns emerged in how pedagogical interaction quality was perceived in Swiss secondary classrooms. The subsequent section discusses key cross-perspective discrepancies across the three domains of teaching—*organization*, *climate*, and *knowledge co-construction*—and offers possible interpretations grounded in theoretical frameworks such as attribution theory (Weiner, 1986), interpersonal theory (Wubbels et al., 2023), the perspective-bound nature of pedagogical practice (Fauth et al., 2020; Helmke and Lenske, 2013), and empirical research on feedback and teaching (Brühwiler and Leutwyler, 2020; Fauth et al., 2020; Wettstein and Scherzinger, 2022). Notably, both the magnitude and direction of these discrepancies varied systematically depending on the specific construct and the overall level of interaction quality in the classroom, suggesting that perspective differences are context-dependent rather than uniform.

Within the domain of classroom *organization*, comparing the two contrast groups—formed based on observer ratings—revealed preliminary patterns in the extent of teacher-student perspective discrepancies, especially regarding teacher *monitoring*. Teachers rated their own monitoring lower than their students did—in both the *High-Interaction Contrast Group (Top-3)* and the *Low-Interaction Contrast Group (Bottom-3)*. This difference in perspective may reflect the distinct roles and interpretive frames of teachers and students: teachers tend to view monitoring as a routine and often subtle form of attention regulation embedded in professional classroom practice. Their assessments might be shaped by idealized standards of classroom management (Helmke, 2009). Students, by contrast, experience the teacher's presence more directly—either as supportive attention or as constant surveillance—and evaluate it in a more situational and less systematic manner (Fauth et al., 2020). This discrepancy was much larger in the *Low-Interaction Contrast Group (Bottom-3)* than in the *High-Interaction Contrast Group (Top-3)*, suggesting that monitoring gaps may widen in more challenging classroom contexts. According to Wubbels' interpersonal model (Wubbels et al., 2023), such a pattern might suggest a disruption of

reciprocity—that is, the degree to which teacher and student behaviors and perceptions align along the dimensions of agency (dominance – submission) and communion (proximity – distance). When reciprocity is disrupted, teacher behavior (e.g., attentive responses) is interpreted differently by students—such as being controlling rather than supportive. In classrooms with higher-rated interaction quality, by contrast, perception and impact appeared more likely to align; interactions seemed more balanced and mutually responsive.

In the teaching domain of *classroom climate*, teachers reported noticeably less aggressive behavior on their part than students attributed to them—in both the *High-Interaction Contrast Group (Top-3)* and the *Low-Interaction Contrast Group (Bottom-3)*. The discrepancy was particularly pronounced in the *Low-Interaction Contrast Group (Bottom-3)*, while it was less pronounced in the *High-Interaction Contrast Group (Top-3)*. One possible explanation for this discrepancy lies in self-serving attributional biases: in stressful situations, teachers may retrospectively externalize or relativize their own aggressive behavior in order to protect their professional self-image (Weiner, 1986). In such cases, they may attribute aggressive episodes not to themselves or their own activities, but to external circumstances such as disruptive student behavior. Through this externalization, the teacher may come to perceive their own behavior as a logical reaction and thus experience it as less aggressive. Students, by contrast, experience such aggressive teacher behavior directly and assign high emotional significance to it—so that even isolated negative experiences can shape their overall judgment (Scherzinger and Wettstein, 2022). These pronounced perspective differences might thus be understood as a result of differences in emotional significance and in self-protective attribution patterns. In addition, the particularly large perspective gap in the *Low-Interaction Contrast Group (Bottom-3)* could be further explained with reference to Wubbels' interpersonal model (Wubbels et al., 2023). In these classrooms, *reciprocity*—understood as the mutual alignment of teacher and student behavior within interpersonal exchanges—appeared disrupted: while the teacher might intend to reassert authority or set boundaries, students might perceive the same behavior as dismissive, demeaning, or aggressive. This divergence can be understood specifically in terms of the interpersonal model's *Communion* dimension, which reflects the degree of perceived *proximity* vs. *distance* in teacher-student relationships. In the *High-Interaction Contrast Group (Top-3)*, by contrast, intention and perception appeared to align more closely, as *reciprocity* seemed to support ongoing, constructive interaction even in challenging situations.

In the teaching domain of *knowledge co-construction*, teachers' and students' assessments of cognitive activation differed considerably—in both the *High-Interaction Contrast Group (Top-3)* and the *Low-Interaction Contrast Group (Bottom-3)*. Teachers consistently evaluated their own behavior regarding cognitive activation more critically than students did, representing the largest perspective discrepancy observed across all constructs ($d' > 1.0$ in the *High-Interaction Contrast Group*). This systematic underestimation may be explained by several interrelated mechanisms. First, due to their expertise and professional knowledge, teachers may set higher standards for teaching quality than students do (Fauth et al., 2020). Students,

by contrast, may evaluate cognitive demands based on their experiences of success or subjective effort without considering the need for in-depth understanding or elaborate mental processes (Fauth et al., 2020). Moreover, teachers' self-assessments of cognitive activation may be shaped by their awareness of the gap between their intentions and actual implementation—what they planned to achieve vs. what they perceive as having been realized in the moment (Helmke, 2009). Students, lacking access to the teacher's planning process, evaluate only the enacted instruction. Notably, unlike *monitoring* or *wellbeing*, this discrepancy remained equally large across both contrast groups, suggesting that it reflects a fundamental role-based perceptual gap rather than being contingent on overall classroom quality.

Across all teaching domains, perspective alignment appeared to vary with the inferential demands of the constructs being assessed. Low-inference constructs such as *structuring*—where assessment requires minimal interpretation—revealed strong convergence across all three perspectives in both contrast groups. In contrast, higher-inference constructs such as *monitoring*, *classroom disturbances*, and *wellbeing*—where assessment requires interpretation of intentions, affective states, or subtle behavioral cues—revealed larger perspective differences that sometimes reversed in direction between contrast groups. This pattern aligns with the distinction between low-inference and high-inference constructs in classroom observation research (Fauth et al., 2020; Praetorius et al., 2014), suggesting that the extent to which teachers, students, and observers align in their perceptions may depend not only on their respective roles and interpretive frames but also on the observable vs. interpretive nature of the construct itself. Constructs requiring greater interpretation appear more susceptible to perspective-dependent assessment, which has implications for both the interpretation of multi-perspective data and the design of teacher feedback.

Third, we explored whether and how the multi-perspective findings can be presented to teachers in ways that are evidence-based and useful for practice. The present pilot study suggests that these findings can be translated into feedback formats that are both accessible and practically relevant for teachers. In particular, the clear overview provided by the radar plot and the direct pairwise perspective comparisons made perceptual differences visible in a way that appeared to support reflection on teaching practice. The reporting approach—emphasizing absolute perspective differences (d) alongside standardized effect sizes (d') and using graphical visualizations—appeared to enhance the presentation of perceptual discrepancies in an accessible and interpretable format. Such graphical feedback formats resonate with findings from teaching and evaluation research, which emphasize that effect sizes and visualizations enhance the interpretability of results (Brühwiler and Leutwyler, 2020; Valentine et al., 2015). However, it is important to note that feedback reports cannot capture the full complexity of classroom interaction and therefore should be seen as starting points for professional dialogue rather than definitive evaluations. The pilot study thus provides preliminary evidence that multi-perspective data can be presented in a way that bridges scientific rigor and practical usability, an approach which will be further developed in the main study.

While the multi-perspective design and the innovative measurement of pedagogical interaction quality are strengths, the present pilot study also entails several limitations that need to be considered in a critical evaluation. First, while pilot-level psychometric diagnostics supported the instrument's feasibility, comprehensive validation remains outstanding. The survey instruments are based on established templates from teaching research and have been systematically adapted to the requirements of the multi-perspective design. Internal consistency estimates (Cronbach's alpha) ranged from acceptable to excellent for most scales (see Table 2), though some scales—particularly from the observer perspective—showed lower values (e.g., *structuring*, $\alpha = .38$). Intraclass correlation coefficients indicated substantial classroom-level variance in student ratings ($ICC(1) = .07-.36$) and reliable differentiation between classrooms ($ICC(2) = .49-.88$), supporting the appropriateness of multilevel analyses. Confirmatory factor analyses (CFA) generally supported the expected unidimensional structure across perspectives (Table 3). While two-level student models benefited from larger individual-level samples, between-level estimates and single-level teacher and observer CFAs remained constrained by the small classroom sample ($n = 16$), requiring cautious interpretation of fit indices. These pilot findings informed targeted item revisions and enhanced observation procedures for the main study. Comprehensive validation—particularly regarding construct validity and measurement invariance across all 18 constructs and perspectives—is reserved for the main study. Second, the representative nature of the sample is limited. The data come from 16 secondary school classes in German-speaking Switzerland. There was no systematic control for regional differences or for variation across school tracks. The sample consisted of around two-thirds of students in the extended requirements track and approximately one-third in the basic requirements track, suggesting a certain structural similarity to the overall population at the descriptive level. The initial trends that emerge here will be examined in more detail in the main study. Third, considering these limitations—particularly the small and not fully representative sample—this pilot study deliberately refrained from using inferential statistical tests. Instead, the analysis focused on absolute perspective differences (Δ), standardized effect sizes (d'), and graphical representations to highlight the practical relevance of perspective differences within specific classes, rather than aiming to generalize the findings from the sample to the population. This approach follows recommendations from evaluation research, according to which effect sizes—especially in exploratory study designs with limited sample sizes—allow for a more meaningful interpretation than significance tests (Lind, 2012; Valentine et al., 2015). Fourth, the ecological validity of the collected data is limited. Although collected in the natural classroom environment, these data represent snapshots that capture the complex dynamics of pedagogical interactions only to a limited extent. Teachers and students were asked to rate how typical the observed lessons were compared to their usual classroom behavior across all three teaching domains; these representativity ratings revealed no systematic deviations. Moreover, the observation data were collected by individual

raters without systematic inter-rater reliability assessment, which limits the objectivity of the external perspective. To address these limitations, video-based observation and coding—including dual or multiple coding procedures with formal inter-rater reliability checks—as well as longitudinal analyses are planned as part of the main study. These will provide a more nuanced understanding of interaction processes and make temporal developments visible (Helmke, 2009; Pauli and Reusser, 2006; Praetorius et al., 2012).

Finally, common method variance and social desirability biases may have affected the self-report data from teachers and students. Teachers may have presented their classroom management and instructional behavior in a more favorable light to maintain a positive professional self-image (Kunter and Baumert, 2006), while students' ratings could be influenced by halo effects stemming from their overall relationship with the teacher (Fauth et al., 2020). The multi-perspective design partially mitigates this concern by triangulating self-reports with external observations; nevertheless, discrepancies between perspectives cannot be attributed solely to perceptual differences but may also reflect measurement artifacts.

In sum, the pilot study demonstrated that pedagogical interaction quality can be meaningfully assessed from multiple perspectives using the INTAKT instrument. Initial psychometric diagnostics—including internal consistency, intraclass correlations, and confirmatory factor analyses—supported the instrument's feasibility, though comprehensive validation is reserved for the main study. The multi-perspective analysis of pedagogical interaction quality reveals that differences in perspective varied depending on both the specific teaching domain and the overall level of pedagogical interaction quality. Some constructs, such as *teacher monitoring*, demonstrated clear dependencies on perspective between the two contrast groups. In these constructs, the magnitude of teacher-student discrepancies shifted systematically with the overall interaction quality: in the *High-Interaction Contrast Group (Top-3)*, teachers' and students' assessments were more aligned, whereas in the *Low-Interaction Contrast Group (Bottom-3)* they increased. This suggests that the perception of these constructs may be sensitive to the overall classroom context. Others, such as *cognitive activation*, appeared to be shaped primarily by role- and process-specific perception patterns. These findings highlight the importance of a differentiated interpretation of multi-perspective data. They also emphasize the need to consider the teaching domain, and the perception patterns of the actors involved when providing teachers with feedback. The feedback reports clearly identified differences in perspective between teachers, students, and observers. These insights can serve as a starting point for reflective discussions about the specific patterns of perception and behavior among teachers and students.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, in anonymized form, upon reasonable request.

Ethics statement

The studies involving humans were approved by Ethikkommission der Pädagogischen Hochschule Bern. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants and/or their legal guardians/next of kin.

Author contributions

LS: Writing – review & editing, Formal analysis, Project administration, Conceptualization, Writing – original draft, Methodology, Data curation, Software, Visualization, Investigation. BE: Investigation, Software, Visualization, Funding acquisition, Formal analysis, Data curation, Methodology, Writing – review & editing, Conceptualization. AW: Writing – review & editing, Conceptualization, Funding acquisition, Supervision. BF: Writing – review & editing, Conceptualization.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by Swiss National Science Foundation, grant number 100019_219487. The APC was funded by the Swiss National Science Foundation.

References

- Arbuckle, C., and Little, E. (2004). Teachers' perceptions and management of disruptive classroom behaviour during the middle years (years five to nine). *Aust. J. Educ. Dev. Psychol.* 4, 59–70.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67 (1), 1–48. doi: 10.18637/jss.v067.i01
- Beaman, R., Wheldall, K., and Kemp, C. (2007). Recent research on troublesome classroom behaviour: a review. *Australas. J. Spec. Educ.* 31, 45–60. doi: 10.1080/10300110701189014
- Bleck, V., and Lipowsky, F. (2019). "Was wissen wir über guten unterricht?—ein update," in *Unterrichtsqualität: Konzepte und Bilanzen Gelingenden Lehrens und Lernens*, eds. R. Messner, and U. Steffens (Münster: Waxmann), 219–249.
- Brunner, E. J., and Noack, P. (2010). "Lehrer-Schüler-Interaktion," in *Handwörterbuch Pädagogische Psychologie*, ed. D. H. Rost (Weinheim: Beltz), 382–389.
- Brühwiler, C., and Leutwyler, B. (2020). Praxisrelevanz von forschung als gemeinsame aufgabe von wissenschaft und praxis: entwurf eines angebots-nutzungs-modells. *Beitr. Lehr. Lehrbild.* 38, 21–36. doi: 10.25656/01:21772
- Can, Y., and Guven, M. (2025). Development of the teacher-student interaction scale. *West. Anatol. J. Educ. Sci.* 16 (1), 1243–1274. doi: 10.51460/baedb.1585671
- Charalambous, C. Y., and Praetorius, A.-K. (2025). Are we doing okay? Making the case for critically examining small and inconsistent effects in research on teaching quality. *Sch. Eff. Sch. Improv.* 36 (2), 153–163. doi: 10.1080/09243453.2025.2482574
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educ. Res.* 38, 181–199. doi: 10.3102/0013189X08331140
- Di Lisio, G., Mila Roa, A., Halty, A., Berastegui, A., Couso Losada, A., and Pitillas, C. (2025). Nurturing bonds that empower learning: a systematic review of the significance of teacher-student relationship in education. *Front. Educ.* 2, 1522997. doi: 10.3389/feduc.2025.1522997
- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., and Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: a moderated mediation analysis of teacher candidates. *J. Educ. Psychol.* 106, 569–583. doi: 10.1037/a0035504
- Doyle, W. (2006). "Ecological approaches to classroom management," in *Handbook of Classroom Management: Research, Practice, and Contemporary Issues*, eds. C. M. Evertson, and C. S. Weinstein (Mahwah, NJ: Lawrence Erlbaum Associates), 97–125.
- Dweck, C. S. (2000). *Self-theories: Their Role in Motivation, Personality, and Development*. New York: Psychology Press. Available online at: <https://www.taylorfrancis.com/books/mono/10.4324/9781315783048/self-theories-carol-dweck> (Accessed March 10, 2025).
- Eckstein, B. (2018). *Gestörter Unterricht: Wie Lehrpersonen, Schülerinnen und Schüler Unterrichtsstörungen wahrnehmen, deuten, bewerten und beschreiben*. [dissertation thesis]. Zurich: University of Zurich. Available online at: <https://www.zora.uzh.ch/id/eprint/158353> (Accessed November 29, 2024).
- Eckstein, B., and Wettstein, A. (2024). Zur qualität pädagogischer interaktionen im unterricht: eine neuartige theoretische konzeptualisierung und eine innovative 2(3 (3-systematik. *Schweiz. Z. Heilpädagogik* 30, 2–9. doi: 10.57161/z2024-03-01
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., and Büttner, G. (2014). Student ratings of teaching quality in primary school: dimensions and prediction of student outcomes. *Learn. Instr.* 29, 1–9. doi: 10.1016/j.learninstruc.2013.07.001
- Fauth, B., Göllner, R., Lenske, G., Praetorius, A.-K., and Wagner, W. (2020). Who sees what? Conceptual considerations on the measurement of teaching quality from different perspectives. *Z. Pädagogik, Beiheft* 66, 138–155. doi: 10.3262/ZPB2001138

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2026.1781315/full#supplementary-material>

- Fauth, B., and Ruth-Herbein, E. (2021). Unterrichtsfeedbackbogen Tiefenstrukturen (UFB). Institut für Bildungsanalysen Baden-Württemberg (IBBW). Available online at: <https://ibbw-bw.de> (Accessed December 11, 2024).
- Fend, H. (1980). *Theorie der Schule*. München: Urban und Schwarzenberg.
- Grewe, N. (2016). "Soziale Interaktion und Klassenklima," in *Lehrer-Schüler-Interaktion*, ed. M. K. W. Schweer (Wiesbaden: Springer), 547–560. doi: 10.1007/978-3-658-15083-9_24
- Hall, J. A., Schmid Mast, M., and West, T. V. (2016). *The Social Psychology of Perceiving Others Accurately*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781316181959
- Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. London: Routledge. Available online at: <https://www.taylorfrancis.com/books/mono/10.4324/9780203181522/visible-learning-teachers-john-hattie> (Accessed March 10, 2025).
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Klett-Kallmeyer.
- Helmke, A., and Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung. *Beitr. Lehr. Lehrbild.* 31, 214–233. doi: 10.25656/01:13848
- Hu, L.-T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 6 (1), 1–55. doi: 10.1080/10705519909540118
- Jang, H., Reeve, J., and Deci, E. L. (2010). Engaging students in learning activities: it is not autonomy support or structure but autonomy support and structure. *J. Educ. Psychol.* 102, 588–600. doi: 10.1037/a0019682
- Klieme, E. (2019). "Unterrichtsqualität," in *Handbuch Schulpädagogik*, eds. M. Haring, C. Rohlf, and M. Gläser-Zikuda (Münster: Waxmann), 393–408.
- Klieme, E., Schümer, G., and Knoll, S. (2001). "Mathematikunterricht in der Sekundarstufe I: aufgabenkultur und unterrichtsgestaltung," in *TIMSS-Impulse für Schule und Unterricht*, ed. Bundesministerium für Bildung und Forschung, eds. E. Klieme, and J. Baumert (Bonn: BMBF), 43–57. Available online at: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2102306 (Accessed December 9, 2024).
- Kokkinos, C. M. (2007). Job stressors, personality and burnout in primary school teachers. *Br. J. Educ. Psychol.* 77, 229–243. doi: 10.1348/000709905X90344
- Kunter, M., and Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learn. Environ. Res.* 9, 231–251. doi: 10.1007/s10984-006-9015-7
- Kunter, M., Pohlmann, B., and Decker, A.-T. (2020). "Lehrkräfte," in *Pädagogische Psychologie*, eds. E. Wild, and J. Möller (Berlin, Heidelberg: Springer), 269–288. doi: 10.1007/978-3-662-61403-7_11
- Lanphen, J., and Wiedenbauer, G. (2021). "Soziale Integration und Kooperation in inklusiven Klassen," in *Handbuch Schulpsychologie: Psychologie für Schule und Unterricht*, eds. K. Seifried, S. Drewes, and M. Hasselhorn (Stuttgart: Kohlhammer), 347–356. doi: 10.17433/9783170397873
- Lind, G. (2012). *Effektstärken: Statistische, Praktische und Theoretische Bedeutsamkeit Empirischer Studien [online Publication]*. Konstanz: Universität Konstanz. Available online at: <http://kops.uni-konstanz.de/handle/123456789/21776> (Accessed December 18, 2024).
- Pauli, C., and Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Z. Pädagogik* 52, 774–798.
- Perrez, M., Huber, G. L., and Geißler, K. A. (2006). "Psychologie der pädagogischen Interaktion," in *Pädagogische Psychologie*, eds. A. Krapp, and B. Weidenmann (Weinheim: Beltz), 269–355.
- Pianta, R. C. (2016). "Teacher-student interactions: measurement, impacts, improvement, and policy," in *Handbook of Social Influences in School Contexts: Social-Emotional, Motivation, and Cognitive Outcomes*, eds. K. R. Wentzel, and G. B. Ramani (New York: Routledge), 243–261.
- Pianta, R. C., and Hamre, B. K. (2009). Classroom processes and positive youth development: conceptualizing, measuring, and improving the capacity of interactions between teachers and students. *New Dir. Youth Dev.* 2009, 33–46. doi: 10.1002/yn.295
- Pianta, R. C., Hamre, B. K., and Allen, J. P. (2012). "Teacher-student relationships and engagement: conceptualizing, measuring, and improving the capacity of classroom interactions," in *Handbook of Research on Student Engagement*, eds. S. L. Christenson, A. L. Reschly, and C. Wylie (Boston, MA: Springer US), 365–386. doi: 10.1007/978-1-4614-2018-7_17
- Pianta, R. C., Hamre, B., and Stuhlman, M. (2003). "Relationships between teachers and children," in *Handbook of Psychology*, ed. I. B. Weiner (Hoboken, NJ: John Wiley & Sons), 199–234. doi: 10.1002/0471264385.wai0710
- Praetorius, A.-K., Klieme, E., Kleickmann, T., Brunner, E., Lindmeier, A., Taut, S., et al. (2020). "Towards developing a theory of generic teaching quality: origin, current status, and necessary next steps regarding the three basic dimensions model," in *Empirische Forschung zu Unterrichtsqualität: Theoretische Grundfragen und Quantitative Modellierungen*, eds. A.-K. Praetorius, T. Kleickmann, and E. Klieme (Weinheim, Basel: Beltz Juventa), 15–36. Available online at: https://www.pedocs.de/frontdoor.php?source_opus=25861 (Accessed October 20, 2024).
- Praetorius, A.-K., Lenske, G., and Helmke, A. (2012). Observer ratings of instructional quality: do they fulfill what they promise? *Learn. Instr.* 22, 387–400. doi: 10.1016/j.learninstruc.2012.03.002
- Praetorius, A.-K., McIntyre, N. A., and Klassen, R. M. (2017). Reactivity effects in video-based classroom research: an investigation using teacher and student questionnaires as well as teacher eye-tracking. *Z. Erziehungswiss* 20, 49–74. doi: 10.1007/s11618-017-0729-3
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., and Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learn. Instr.* 31, 2–12. doi: 10.1016/j.learninstruc.2013.12.002
- R Core Team. (2026). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (Accessed January 5, 2026).
- Revelle, W. (2025). psych: Procedures for psychological, psychometric, and personality research (Version 2.5.6) [Computer software]. Northwestern University. Available online at: <https://CRAN.R-project.org/package=psych> (Accessed January 5, 2026).
- Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48 (2), 1–36. doi: 10.18637/jss.v048.i02
- Scherzinger, M., and Wettstein, A. (2019). Classroom disruptions, the teacher-student relationship and classroom management from the perspective of teachers, students and external observers: a multimethod approach. *Learn. Environ. Res.* 22, 101–116. doi: 10.1007/s10984-018-9269-x
- Scherzinger, M., and Wettstein, A. (2022). *Beziehungen in der Schule Gestalten*. Stuttgart: Kohlhammer.
- Schwarzwälder, L., and Lohse-Bossenz, H. (2021). "Professionelle Beziehungswahrnehmung als zentrale Voraussetzung für die Entstehung und Gestaltung von sozialen Beziehungen im Kontext von Schule und Unterricht," in *Soziale Eingebundenheit. Sozialbeziehungen im Fokus von Schule und LehrerInnenbildung*, eds. G. Hagenauer, and D. Raufelder (Münster: Waxmann), 85–98. doi: 10.31244/9783830992660
- Schweer, M. K. W., Thies, B., and Lachner, R. P. (2017). "Soziale Wahrnehmungsprozesse und unterrichtliches Handeln: eine dynamisch-transaktionale Perspektive," in *Lehrer-Schüler-Interaktion*, ed. M. K. W. Schweer (Wiesbaden: Springer), 121–145. doi: 10.1007/978-3-658-15083-9_5
- Seidel, T. (2015). "Klassenführung," in *Pädagogische Psychologie*, eds. E. Wild, and J. Möller (Berlin, Heidelberg: Springer), 107–119. doi: 10.1007/978-3-642-41291-2_5
- Siegler, R., Eisenberg, N., DeLoache, J., and Saffran, J. (2016). "Moralentwicklung," in *Entwicklungspsychologie im Kindes- und Jugendalter*, eds. R. Siegler, N. Eisenberg, J. DeLoache, J. Saffran, and S. Pauen (Berlin, Heidelberg: Springer), 529–573. doi: 10.1007/978-3-662-47028-2_14
- Tabachnick, B. G., and Fidell, L. S. (2014). *Using Multivariate Statistics*. 6th ed. Harlow, UK: Pearson.
- Thiel, F. (2016). *Interaktion im Unterricht: Ordnungsmechanismen und Störungsdynamiken*. Opladen, Toronto: Verlag Barbara Budrich. doi: 10.36198/9783838545714
- Thies, B. (2017). "Forschungszugänge zur Lehrer-Schüler-Interaktion," in *Lehrer-Schüler-Interaktion*, ed. M. K. W. Schweer (Wiesbaden: Springer), 65–88. doi: 10.1007/978-3-658-15083-9_3
- Timmermans, A. C., Rubie-Davies, C. M., and Wang, S. (2021). Adjusting expectations or maintaining first impressions? The stability of teachers' expectations of students' mathematics achievement. *Learn. Instr.* 75, 101483. doi: 10.1016/j.learninstruc.2021.101483
- Valentine, J. C., Aloe, A. M., and Lau, T. S. (2015). Life after NHST: how to describe your data without "p-ing" everywhere. *Basic Appl. Soc. Psychol.* 37, 260–273. doi: 10.1080/01973533.2015.1060240
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M., and Pietsch, M. (2020). "Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts: ein kritischer Vergleich verschiedener Modellvarianten," *Empirische Forschung zu Unterrichtsqualität - Theoretische Grundfragen und Quantitative Modellierungen*, eds. A.-K. Praetorius, T. Kleickmann, and E. Klieme (Münster: Waxmann), 63–80. doi: 10.25656/01:25864
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., and Trautwein, U. (2016). Student and teacher ratings of instructional quality: consistency of ratings over time, agreement, and predictive power. *J. Educ. Psychol.* 108, 705–721. doi: 10.1037/edu0000075
- Wang, M.-T., Hofkens, T., and Ye, F. (2020). Classroom quality and adolescent student engagement and performance in mathematics: a multi-method and multi-informant approach. *J. Youth Adolesc.* 49 (9), 1987–2002. doi: 10.1007/s10964-020-01195-0
- Weiner, B. (1986). *An Attributional Theory of Motivation and Emotion*. New York, NY: Springer US. doi: 10.1007/978-1-4612-4948-1
- Wemmer-Rogh, W., Gossner, L., Wehrli, F., and Praetorius, A.-K. (2023). INSULA 2.0: Instrumentarium zur Unterrichtsbeurteilung ausgerichtet auf den Lehrplan 21 im Auftrag von argev. Zenodo. doi: 10.5281/zenodo.8280334

- Werth, L., Denzler, M., and Mayer, J. (2020). "Soziale wahrnehmung." *Sozialpsychologie – Das Individuum im Sozialen Kontext*, L. Werth, M. Denzler, and J. Mayer (Berlin, Heidelberg: Springer), 137–186. doi: 10.1007/978-3-662-53897-5_5
- Wettstein, A. (2008). *Beziehungsgestaltung und Gewaltprävention in der Schule: Das BASYS-Konzept zur Förderung von Beziehungs- und Sozialkompetenz bei Lehrkräften*. Bern: Haupt.
- Wettstein, A. (2012). A conceptual frame model for the analysis of aggression in social interactions. *J. Soc. Evol. Cult. Psychol.* 6, 141–157. doi: 10.1037/h0099218
- Wettstein, A. (2013). Die wahrnehmung sozialer prozesse im unterricht. *Schweiz. Z. Heilpädagogik* 19, 5–13.
- Wettstein, A., and Scherzinger, M. (2022). *Unterrichtsstörungen Verstehen und Wirksam Vorbeugen*. 2nd ed. Stuttgart: Kohlhammer.
- Wettstein, A., Scherzinger, M., and Ramseier, E. (2018). Empirische arbeit: unterrichtsstörungen, beziehung und klassenführung aus lehrer-, schüler- und beobachterperspektive. *Psychol. Erzieh. Unterr* 65 (1), 58–74. doi: 10.2378/peu2018.art04d
- Wettstein, A., Scherzinger, M., and Wyler, S. (2016). Fragebogen zur erfassung von störungen im unterricht: die faktorenstruktur der schülerinnen- und schülerversion. *Emp. Sonderpädagogik* 8 (2), 189–202. urn:nbn:de:0111-pedocs-123045.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4 (43), 1686. doi: 10.21105/joss.01686
- Wubbels, T., Brekelmans, M., Den Brok, P., Wijsman, L., Mainhard, T., and Van Tartwijk, J. (2015). "Teacher-student relationships and classroom management," in *Handbook of Classroom Management*, 2nd ed. eds. E. T. Emmer, and E. J. Sabornie (New York: Routledge), 363–386.
- Wubbels, T., Brekelmans, M., and Hooymayers, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teach. Teach. Educ.* 8, 47–58. doi: 10.1016/0742-051X(92)90039-6
- Wubbels, T., Mainhard, T., Den Brok, P., Claessens, L., and Van Tartwijk, J. (2023). "Classroom management at different timescales," *Handbook of Classroom Management*, E. J. Sabornie, and D. L. Espelage (New York: Routledge), 388–414. doi: 10.4324/9781003275312-25