

DOI: <https://doi.org/10.57125/FED.2026.09.10>

How to cite: Ritt, C. (2026). What do students really judge? IVET teachers' perspectives on student evaluation as a source of professional assessment. *Futurity Education*, 6(3), 162–174.
<https://doi.org/10.57125/FED.2026.09.10>

What Do Students Really Judge? IVET Teachers' Perspectives on Student Evaluation as a Source of Professional Assessment

Cinzia Ritt

Department of Philosophy and Human Sciences, Cà Foscari University of Venice, Italy

<https://orcid.org/0009-0008-6470-0512>

Corresponding author: cinzia.ritt@unive.it.

Received: May 17, 2026 | **Accepted:** August 15, 2026 | **Available online:** September 3, 2026

Abstract: Student evaluation of teaching remains a widely used but theoretically contested institutional practice, with ongoing debate over whether student judgement can be interpreted as a valid indicator of teaching quality. This debate remains largely unexamined for teachers in initial vocational education and training, a professional category that, within the Italian regional vocational training system, lacks the standardised qualification pathways applicable to school-based teachers. Drawing on a capability-based theoretical framework, this study examined how vocational teachers position student evaluation within their broader understanding of professional assessment and proposed an alternative theoretical interpretation of student satisfaction data. Eighteen semi-structured interviews were conducted with vocational teachers employed by a regional training provider in north-eastern Italy, using thematic analysis with a critical-incident technique. Twelve interviews addressed student evaluation directly. Teachers accepted student judgement only conditionally — tied to track record, students' maturity, or explicit openness to unfavourable scores — while questioning its representativeness and its vulnerability to relational and episodic contamination. Procedural conditions of data collection, and opacity in how collected data were subsequently used, shaped whether student judgement was considered sincere and useful. Several teachers proposed outcome-based alternatives to direct opinion, while others grounded

legitimacy in students' continuous presence during teaching. This legitimacy was concentrated in relational and contextual dimensions of practice rather than in technical competence. These findings offer a theoretically grounded account of student evaluation for an unregulated professional category, with practical implications for instrument design in comparable settings.

Keywords: student evaluation of teaching, student satisfaction survey, vocational teachers, professional competence, conversion factors, professional development.

Introduction

Research Problem

Ensuring the quality of vocational education and training (VET) depends, to a significant extent, on the professional development and evaluation of the teachers and trainers who deliver it. Persistent workforce shortages, an ageing teaching population, and concerns about the profession's attractiveness have placed teacher evaluation and continuing professional development (CPD) firmly on the EU VET policy agenda. Psifidou et al. (2026) show, based on pilot evidence from Cedefop's European Vocational Teacher Survey, that VET teachers' professional development remains more closely tied to informal learning and organisational support than to structured evaluation or formal training. The 2025 Herning Declaration explicitly calls for greater professional autonomy and CPD support for VET trainers as a distinct professional group, while Antera (2021) argues that professional competence in vocational teaching should itself be understood as situated, developmental and dynamic, rather than as a fixed, formally certified attribute, a characterisation echoed in recent large-scale evidence that VET teachers themselves treat assessment-related competence as a multilayered, context-dependent construct rather than a uniform skill (Meijer et al., 2025) and further reflected in a dual professionalism through which vocational teachers often navigate an unresolved tension between occupational and pedagogical identity (Bükki & Fehérvári, 2024).

Within this broader evaluation landscape, student evaluation of teaching (SET) is a widely used, if contested, component. Spooren et al. (2013) and Marsh and Roche (1997) document their extensive use in general and higher education, though comparable evidence specific to VET remains scarce; within the case examined in this study, SET data are formally combined with student grades into a composite institutional threshold for VET teacher and trainer evaluation. Spooren et al. (2013) and Uttl et al. (2017) argue that such instruments are often treated as a direct, if imperfect, proxy for teaching quality, without a clear theoretical account of what the data actually capture about a teacher's professional competence—a gap this study addresses directly.

This gap is particularly salient in the Italian regional VET system (Istruzione e Formazione Professionale, IeFP), where teachers and trainers operate under conditions structurally different from those of school-based teachers. Hippe et al. (2022) show that the same occupational title covers substantially different professional configurations across national systems, and recommend defining VET teachers and trainers by function and task rather than by title alone. Antera (2023) further finds that competence acquisition among vocational teachers varies systematically with institutional and biographical conditions of practice, rather than following a single certified pathway. While Law 79/2022 established national requirements for the initial training of school teachers, no equivalent standard currently applies to teachers and trainers employed within regional vocational training providers (Cedefop, 2025; Cedefop, 2023). This regulatory asymmetry leaves VET teachers' evaluation without a shared professional or theoretical reference point.

Research Focus

This study departs from the question of whether student judgement validly measures teaching effectiveness — a question the SET literature has arguably already answered in the negative when effectiveness is defined in terms of learning outcomes (Uttl et al., 2017). Instead, drawing on the Capability Approach (Sen, 1999) and on accounts of professional competence as embedded in situated action (Pastré et al., 2006; Schön, 1983), it advances a different hypothesis: that student evaluation data may be more usefully understood not as a direct measure of a VET teacher's competence, but as a lateral, indirect form of epistemic access to the conversion factors — the contextual conditions of practice — under which that competence is exercised. On this reading, such data would inform professional development trajectories rather than function as a summative judgement of competence itself.

Research Aim and Research Questions

The following research question guides the study: how do IVET teachers employed by ENAIP Veneto, a regional vocational training provider, position the role of student evaluation within their broader account of what should be assessed in their professional work? Addressing this question contributes to a theoretically grounded understanding of evaluation practice in a professional category that remains, at present, without either a shared regulatory standard or an established evaluative framework of its own.

Literature Review

Teacher evaluation has historically been organised around two distinct purposes that frequently pull in different directions: a summative, accountability-oriented function, intended to assure and certify the quality of teaching, and a formative function, oriented toward professional development and improvement (Isoré, 2009), a tension confirmed in recent empirical work on how teacher competence assessment is designed and valued in practice (González-Fernández et al., 2024). Across EU Member States, this tension has recently been reframed in explicitly workforce-oriented terms: persistent shortages, an ageing teaching population, and concerns over the attractiveness of the profession have pushed evaluation and continuing professional development (CPD) closer together in policy discourse, with recent pilot evidence indicating that professional development remains more strongly tied to informal learning and organisational support than to structured evaluation or formal training (Psifidou et al., 2026), and with national policy monitoring pointing to a persistent lack of standardised qualification pathways for teachers outside the school system (Cedefop, 2025). At EU level, this shift has been formalised through explicit calls for greater professional autonomy and CPD support for VET teachers and trainers as a distinct occupational group (Herning Declaration, 2025). In initial vocational education and training (IVET) specifically, this accountability–development tension is compounded by the absence of a stable, standardised professional profile: competence in vocational teaching is increasingly characterised as situated, developmental and dynamic, varying systematically with institutional and biographical conditions of practice rather than following a single certified pathway (Antera, 2021, 2023).

Within this broader evaluation landscape, student evaluation of teaching (SET) occupies a contested position. A first strand of research treats SET as a legitimate, if partial, indicator of teaching quality, best interpreted not as a stand-alone score but as one dimension of a broader, multidimensional teaching profile, most informative when aggregated at class level and combined with structured pedagogical feedback (Marsh & Roche, 1997); recent qualitative analysis of open-ended student comments similarly finds that judgement differentiates along specific dimensions — clarity, assessment alignment, relevance, and support for student success — rather than reducing to a single overall impression (Caldwell et al., 2024). A second, methodologically oriented strand raises substantial doubts about this use: meta-analytic evidence finds no reliable association between SET ratings and independent measures of student learning, the comparability of SET scores across courses, instructors and years has been challenged on psychometric

grounds, and methodological work has highlighted the conditions and limitations that must be considered when interpreting SET data in the first place (Uttl et al., 2017; Stark & Freishtat, 2014; Royal, 2017), a pattern replicated in more recent evidence finding no association between such data and independent, objective measures of student learning (Gilbert & Gilbert, 2025), with instrument-design work explicitly identifying the conflation of consumer satisfaction with teaching quality as a core measurement problem requiring redesign (Dyer & Donnelly-Hermosillo, 2024), and a substantial body of evidence further documenting that such ratings are systematically biased by student and instructor demographics independently of teaching quality (Heffernan, 2022; Kim et al., 2024). A third strand converges on an integrative and contextualising response to this disagreement, recommending that SET should not be used in isolation but combined with classroom observation and competence-based evidence. That generic evaluative items — such as overall teaching effectiveness — be replaced with dimension-specific ones capable of capturing the contextual conditions of practice (Spooren et al., 2013), a recommendation reinforced by recent meta-analytic evidence that student feedback interventions produce only modest gains in teaching quality overall, with larger effects specifically when teachers are supported in interpreting the feedback and encouraged to discuss it with students (Röhl et al., 2025). Across these three strands, the debate on student evaluation has concentrated almost exclusively on its measurement validity as an indicator of teaching quality. Comparatively little attention has been paid to a more basic, prior question, independent of validity: what kind of information SET data actually encode about the professional practice being evaluated.

A separate body of work in activity analysis, professional didactics and reflective-practice theory has challenged the view of professional competence as a stable repertoire of knowledge and skills, proposing instead that competence takes shape in and through situated action — in interpretation, judgement, and the ongoing regulation of activity in concrete professional situations, rather than as a possession that precedes and is subsequently applied to practice (Pastré et al., 2006; Schön, 1983, Le Boterf, 2008). A parallel, largely independent tradition in welfare and development economics argues that the possession of resources or skills is insufficient to guarantee their effective exercise: personal and contextual conditions — conversion factors — determine whether a given resource translates into a real, exercisable capability, and, in turn, into an actual functioning (Sen, 1999; Nussbaum, 2011), a distinction with direct empirical purchase in recent professional-learning research showing that, unlike knowledge and skills, capabilities are demonstrated only in context and resist standardised assessment (Holdsworth et al., 2025). Recent conceptual work specific to vocational teaching converges with both traditions, characterising the professional competence of VET teachers as situated, developmental and variable across institutional and biographical conditions, rather than as a fixed, formally certifiable attribute (Antera, 2021, 2023). Taken together, this literature implies that no single, decontextualised indicator — including a satisfaction score — can be expected to capture professional competence directly, since competence is, on this account, constitutively inseparable from the conditions under which it is exercised.

These two strands of literature rarely intersect directly. Research on student evaluation has almost uniformly asked whether such data validly measure teaching quality or effectiveness; research on situated professional competence has, independently, established that competence cannot be reduced to a decontextualised score in the first place. Read together, they support a shift from a question of measurement validity to one of epistemic status: not whether student judgement can observe competence directly, but what kind of informational content about the conditions of professional action might be embedded in that judgement. This is not a claim already established in either literature, but a possible alternative interpretation of student evaluation data proposed here: rather than treating student judgement as a direct measure of competence, it may provide indirect information about some of the contextual conditions that can be conceptualised, from a capability perspective, as conversion factors, through which competence is exercised, given that students occupy a structurally unique position as

continuous witnesses to the situated conditions of teaching, independently of their capacity to evaluate its technical or disciplinary content. This reframing appears to have received comparatively little systematic attention within the student evaluation literature, which remains oriented largely toward the validity question.

Taken together, the literature identifies three unresolved issues. First, research on teacher evaluation research has yet to establish a widely shared framework for assessing IVET teachers as a distinct professional group, whose competence is increasingly theorised as situated rather than standardised. Second, the debate on student evaluation has focused almost exclusively on the psychometric validity of such data as a measure of teaching quality, rather than on the epistemic status of the information student judgement actually encodes. Third, theories of professional competence consistently conceptualise competence as situated and context-dependent, yet this perspective has rarely informed the interpretation of student evaluation data specifically. The present study addresses these gaps by examining how IVET teachers themselves conceptualise the role of student evaluation within their broader understanding of professional assessment.

Materials and Methods

This study adopts a qualitative research design (Creswell, 2014) to explore how IVET teachers at ENAIP Veneto position themselves toward the evaluation of their professional work, with particular attention to the role attributed to student judgement.

Sample and Participants

Eighteen teachers employed at ENAIP Veneto were interviewed. Participants were selected through purposive sampling following a maximum variation strategy (Patton, 2002), aimed at capturing heterogeneity across three dimensions: subject area (technical/professional disciplines, $n = 10$; cultural/general disciplines, $n = 8$), province of employment (Padova, $n = 9$; Belluno, $n = 3$; Treviso, $n = 3$; Vicenza, $n = 1$; Venezia, $n = 1$), and seniority in the role, ranging from teachers in their first year of employment (school year 2025/2026) to a participant employed since 1990/1991. The sample was gender-balanced (9 women, 9 men), with birth years ranging from 1967 to 1999.

Instrument and Procedure

Data were collected through semi-structured, one-to-one interviews lasting approximately one hour, centred on teachers' professional competencies and agency. Interview questions were developed using the critical incident technique (Flanagan, 1954), prompting participants to ground their reflections in specific, concretely recalled episodes of professional practice rather than general opinions. Interviews were audio-recorded with participants' consent and transcribed using automatic transcription software, with subsequent manual review. The present analysis focuses on participants' responses to the interview's closing question: *"Who would you like to be evaluated by for your work? What aspects should someone evaluating an IVET teacher consider?"*. This question was analysed as part of the broader interview in which it was embedded, rather than in isolation, in order to preserve its interpretive continuity with the preceding discussion of competencies and agency.

Data Analysis

Interview transcripts were analysed using thematic analysis following the six-phase procedure outlined by Braun and Clarke (2006): familiarisation with the data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. Coding proceeded inductively, allowing codes to emerge from the data rather than being imposed from a predefined framework. An initial codebook of 66 codes was generated across the full set of responses to the closing

question. Consistent with Braun and Clarke's (2006) distinction between a rich description of an entire dataset and a detailed account of one particular aspect of it, the analysis was subsequently narrowed to the subset of codes directly related to teachers' positioning toward student evaluation, present in twelve of the eighteen interviews. This subset was organised into seven themes, presented in the Results section.

Results

Analysis of the eighteen interviews identified seven recurring themes concerning teachers' positioning toward student evaluation, present in twelve of the eighteen cases (Table 1). The remaining six participants discussed evaluation broadly without referring to student judgement as a distinct source, a pattern addressed further in the Discussion section.

Across the twelve relevant interviews, student judgement was rarely treated as a self-evidently valid or invalid instrument. Instead, participants actively constructed the conditions under which it could be trusted, contested its representativeness, identified sources of distortion, or proposed indirect alternatives grounded in student outcomes rather than student opinion.

Acceptance of student judgement was the most frequently reported position, though rarely unconditional (Table 1, Theme 1). Trust was grounded differently across accounts: as a general disposition not further justified, as contingent on students' developmental stage, as supported by a consistent track record over time, or as an explicit willingness to accept an unfavourable score without contesting it.

A minority position (Theme 2) questioned the representativeness of student judgement on structural grounds, linking it to features of the student population rather than to the evaluation instrument itself. Unlike Theme 1, this critique did not include a proposed remedy, and one participant reported divergent views among colleagues, suggesting a lack of a shared professional stance within the organisation.

Several accounts (Theme 3) identified specific conditions under which student judgement could be distorted independently of teaching quality: interpersonal conflict, illustrated by one participant through a hypothetical example of a lowered score following a disagreement with a student, and the timing of data collection relative to classroom events such as assessments. A related concern, that judgement formed on a single episode may not represent an extended teaching relationship, was raised independently of the conflict-based account, suggesting two distinct sources of concern about representativeness at the level of individual data points rather than of the student population as a whole (Theme 2).

Two participants named procedural features of data collection as relevant to its value, though pointing in different directions (Theme 4). Anonymity in data collection was described as a precondition for sincerity. Conversely, a requirement for students to justify scores below a given threshold was described as valuable, enabling teachers to understand and act on the reasons behind a low score rather than simply receiving it.

A related but distinct concern addressed what happens to data once collected (Theme 5): two participants reported that evaluation outcomes, whether an individual score or aggregated gradimento data, were not returned to or discussed with the person evaluated. Both participants described this independently: one referred to an individual score and the other to aggregate satisfaction data, suggesting the concern is not limited to a single evaluation instrument.

A further set of proposals (Theme 6) shifted the evaluative focus away from direct student opinion toward student-associated outcomes: end-of-programme satisfaction, a learning-gain comparison between starting and ending levels, aggregate pass-rate indicators, and internship outcomes. These proposals were framed as complements to, rather than replacements for, direct student judgement in three of the four accounts.

Finally, one account (Theme 7) grounded the legitimacy of student judgement in students' direct co-presence during the teaching event, describing them as the only available witnesses to what occurs once the classroom door is closed. A related emphasis on the content transmitted, rather than the teacher's likeability, as the basis of student judgement was independently reported in the context of Theme 1.

Table 1

Themes concerning IVET teachers' positioning toward student evaluation (n = 18 interviews)

Theme	Description	Interviews (n)	Illustrative excerpt
1. Conditional acceptance	Student judgement accepted, contingent on trust, maturity, track record, or openness to negative scores	ID01, ID05, ID12, ID14, ID15 (n=5)	"They don't evaluate me for being likeable, but for what I actually convey" (ID12)
2. Structural critique of representativeness	Judgement questioned on structural grounds; divergent views reported among colleagues	ID06, ID18 (n=2)	"Many come here because they must complete three years... it isn't a fair measure" (ID06)
3. Contamination by relational/episodic factors	Vulnerability to personal conflict (illustrated hypothetically) or to assessment timing	ID01, ID04, ID15, ID18 (n=4)	"If I argued with the teacher that morning, the evaluation is already off" (ID15, on timing)
4. Procedural conditions of reliability	Anonymity for sincerity vs. justification requirement for constructive follow-up	ID12, ID15 (n=2)	"It should be explained to students that a score below 6 needs a reason" (ID15)
5. Opacity of collected data	Data collected but not returned or discussed	ID03, ID15 (n=2)	"I asked for my score the previous year; it was not given to me" (ID03)
6. Alternatives to direct judgement	Learning gain, satisfaction, internship and pass-rate outcomes as complements	ID06, ID07, ID08, ID16 (n=4)	"It should be based on how much students improve" (ID16)
7. Epistemic legitimacy of judgement	Legitimacy grounded in co-presence during teaching	ID03 (n=1)	"It's them and me behind that door" (ID03)

Note: n indicates the number of interviews in which the theme was present out of eighteen total.

Source: prepared based on the study's interview data.

Discussion

Twelve of the eighteen interviews addressed student evaluation directly; the remaining six discussed professional evaluation more broadly without treating student judgement as a distinct source. This distribution is itself informative: when asked what should be considered in evaluating their professional work, student satisfaction surveys were not a universally salient reference point for teachers. Where it did surface, however, it emerged as a source teachers actively condition, qualify and situate, rather than one they treat as self-evidently valid or invalid — a pattern that provides empirical grounding for this study's central theoretical proposition.

Conditional acceptance (Theme 1) is the clearest illustration. None of the accounts endorsing student judgement did so unconditionally; trust was consistently tied to something external to the judgement itself — a track record, a developmental threshold, procedural safeguards. This mirrors, from the institutional side of evaluation, recent evidence that department chairs and directors interpreting student feedback for personnel decisions employ comparable contextualising strategies — disregarding outliers, weighing response rates, seeking corroborating evidence — rather than treating the raw data as self-sufficient (Drue & Bifulco, 2025). This sits awkwardly with either pole of the existing SET debate, which asks whether

student judgement is a valid indicator of quality (Marsh & Roche, 1997) or is not (Uttl et al., 2017). Participants were, in effect, answering a different question than the one the psychometric literature has posed: not whether the judgement is valid, but under what conditions it becomes informative.

The procedural conditions (Theme 4) and opacity (Theme 5) themes provide direct, contextually specific confirmation of concerns raised in the methodological SET literature. The critique that generic, decontextualised items fail to capture the conditions of practice (Spooren et al., 2013) is echoed almost verbatim in participants' objections to specific features of ENAIP Veneto's existing student satisfaction questionnaire — the absence of anonymity guarantees, the requirement to justify low scores, the non-return of aggregated results to those evaluated. Recent work reframing SET as an occasion for dialogue and reflection on teaching, rather than as a measurement exercise in itself, offers a useful lens here (Søndergaard et al., 2024): several teachers' objections were not to being evaluated by students as such, but to a procedure that forecloses exactly the kind of dialogic use this literature recommends.

The structural critique (Theme 2) and contamination (Theme 3) themes might, from a purely psychometric standpoint, be read as further confirmation of validity concerns already well documented (Stark & Freishtat, 2014; Uttl et al., 2017), with recent systematic evidence continuing to locate students at the centre of measurement inconsistencies documented in the field (Quansah et al., 2024). A recent four-level analytic model likewise identifies context and timing as recurrent sources of inconsistency (Cook et al., 2024). Dedicated methodological work confirms that timing relative to the final examination systematically shapes SET outcomes (Vehovar & Štrlekar, 2024)—directly echoing participants' concern that the timing of data collection could influence judgement independently of teaching quality. This concern is not, however, uniformly supported: large-scale analysis has found several commonly feared distorting factors — class size, student demographics, assessment timing — to have negligible systematic effect (O'Donovan, 2024), suggesting the salience of these conditions for formatori may reflect situated professional concern rather than documented empirical regularity. Course design, moreover, has been shown to predict student-perceived quality and engagement more strongly than characteristics attributed to the individual teacher (Levinsson et al., 2024) — a pattern broadly consistent with a conversion-factors reading of the present findings, though interpretive caution about what such inconsistencies license remains an established methodological concern (Royal, 2017). The interpretation proposed here nonetheless differs from a purely psychometric reading: treating student judgement as a lateral form of epistemic access rather than a direct measure of competence turns what the validity literature treats as noise — susceptibility to relational conflict, cohort composition, timing — into part of the signal, evidence of the conversion factors the judgement is proposed to index, rather than random measurement error. This reframing is offered as an interpretive proposal grounded in the data, not as an already-established finding.

The alternatives theme (Theme 6) is notable for a different reason: several participants, independently and without prompting toward any particular theoretical framework, proposed measures — learning gain relative to starting point, internship outcomes, pass-rate indicators — that shift the evaluative object away from raw student opinion toward outcomes shaped by context. This spontaneous move mirrors, from the practitioner's side, the same reframing the multidimensional-profile literature proposes from the measurement side (Marsh & Roche, 1997).

The epistemic legitimacy theme (Theme 7) offers the most direct empirical basis for the study's central claim, and it is worth being precise about the direction of inference: the theoretical proposal advanced in this paper was not imposed on the data in advance, but is substantially anticipated by participants' own reasoning. The justification that students are legitimate judges specifically because they are the only witnesses present once the classroom door is closed deserves unpacking rather than simple assertion. Students are continuously present throughout the delivery of teaching. They are therefore positioned to observe repeated patterns of interaction and to directly experience relational and

organisational dimensions of practice that other potential evaluators typically encounter only intermittently. This continuous presence, however, does not extend equally to every dimension of professional practice: the same participants who grounded legitimacy in co-presence also emphasised that what students judge is *content transmitted*, not the teacher's technical or disciplinary competence as such. Read together, these findings suggest that the epistemic value of student judgement may be dimension-specific rather than global — informative regarding the relational and contextual conditions of practice that students witness directly, considerably less so regarding technical or disciplinary competence that lies largely outside their observational access. This is a more precise claim than asking whether SET is valid or invalid overall, and it aligns with recent instrument-design work explicitly built around the same principle—distinguishing what students can be expected to judge validly from what they cannot (Woodzicka et al., 2025). This relational basis of legitimacy also resonates with broader evidence that learner-centred relational qualities — warmth, empathy, responsiveness — are consistently associated with positive student outcomes (Cornelius-White, 2007), lending independent support to the weight participants placed on the relational dimension of practice as something students are well positioned to judge.

Read against comparative evidence, these findings also speak to a broader, cross-national pattern documented in VET research: the same occupational title conceals substantially different professional configurations depending on institutional and biographical conditions (Antera, 2021, 2023; Hippe et al., 2022), and professional development across European VET systems is more closely tied to informal, contextual conditions than to standardised, formal instruments (Psifidou et al., 2026). The present findings extend this pattern to evaluation specifically: just as competence itself resists standardised description, the data suggest that the informational content of the instruments used to evaluate it does too.

Limitations

These findings should be read with several limitations in mind. The sample is drawn from a single regional VET provider, and the idiographic depth this affords comes at the cost of generalisability beyond the Italian IeFP context. The twelve interviews addressing student evaluation directly are a subset of a broader, purposively — not statistically — sampled corpus, and the six teachers who did not raise student judgement as a distinct source cannot be assumed to hold no view on the matter; the interview question did not prompt for it directly. The cross-sectional, retrospective nature of the interviews does not allow claims about how these positions might change over time or in response to a redesigned instrument.

Conclusions

This study examined how IVET teachers employed by a regional Italian vocational training provider position student evaluation within their broader account of professional assessment. Rather than asking whether student satisfaction data validly measures teaching competence — a question existing literature has largely already answered in the negative when competence is equated with learning outcomes — the study asked a different, prior question: what conditions IVET teachers themselves attach to the legitimacy and usefulness of student judgement.

Seven themes emerged from the twelve interviews in which student evaluation was addressed directly, out of eighteen conducted in total: conditional acceptance, structural critique of representativeness, contamination by relational and episodic factors, procedural conditions of reliability, opacity in the use of collected data, alternatives grounded in student outcomes rather than opinion, and legitimacy grounded in students' epistemic position as continuous witnesses to practice. Read together, these themes indicate that teachers do not treat student judgement as a self-evidently valid or invalid instrument, but as one whose informational value depends on the conditions under which it is produced, collected and used.

This pattern lends empirical support to the study's central theoretical proposition: that student satisfaction survey data may be more usefully understood as a lateral, indirect form of epistemic access to the conditions of practice — the conversion factors, in Capability Approach terms — under which competence is exercised, rather than as a direct measure of competence itself. The findings further suggest that this epistemic value is not uniform across all dimensions of practice, but concentrated in the relational and contextual aspects of teaching that students are structurally positioned to observe, and considerably weaker with respect to technical or disciplinary competence that lies largely outside their observational access.

These findings carry practical relevance for ENAIP Veneto and, tentatively, for IVET providers operating under comparable conditions of institutional fragmentation. Several features of the existing gradimento instrument identified by participants as problematic—the absence of anonymity guarantees, the requirement to justify low scores, the non-return of aggregated data—are, in principle, addressable through instrument redesign rather than by abandoning student evaluation altogether. More broadly, the findings suggest that evaluation frameworks for IVET teachers might usefully distinguish between what a given evaluative source can be expected to reveal about the conditions of practice and what it cannot, rather than treating any single source as a comprehensive measure of professional competence.

Suggestions for Future Research

Future studies could test, across more structurally comparable IVET contexts, whether the conditions identified here as shaping the perceived credibility of student judgement — anonymity, procedural transparency, outcome-based framing — are empirically associated with teachers' willingness to treat student evaluation data as informative. A second direction would extend the comparative lens already applied by cross-national VET research to the question of evaluation specifically, alongside the question of competence with which that research has so far been more directly concerned. A third, more targeted direction follows from the dimension-specific reading of epistemic access proposed here: future instruments could test empirically which aspects of professional practice student judgement tracks reliably, and which it does not.

Acknowledgements

The author thanks ENAIP Veneto for co-funding the doctoral fellowship within which this research was conducted, and the eighteen teachers who participated in the study.

Conflict of Interest

None.

Funding

The Author received no funding for this research.

References

- Antera, S. (2021). Professional competence of vocational teachers: A conceptual review. *Vocations and Learning*, 14(3), 459–479. <https://doi.org/10.1007/s12186-021-09271-7>
- Antera, S. (2023). Competence importance and acquisition: Comparing qualified and non-qualified vocational teachers. *Journal of Education and Work*, 36(2), 109–124. <https://doi.org/10.1080/13639080.2023.2167956>

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bükki, E., & Fehérvári, A. (2024). Teacher, professional or both? A mixed method study of the professional identity of vocational teachers and trainers in Hungary. *International Journal for Research in Vocational Education and Training*, 11(3), 396–428. <https://doi.org/10.13152/IJRVET.11.3.4>
- Caldwell, D., Johnson, C., Moore, M., Moore, A., Poush, M., & Franks, A. M. (2024). Teaching through the student lens: Qualitative exploration of student evaluations of teaching. *American Journal of Pharmaceutical Education*, 88(3), Article 100672. <https://doi.org/10.1016/j.ajpe.2024.100672>
- Cedefop. (2023, March 2). *Italy: Approval of national implementation plan (VET)* [Policy timeline entry]. <https://www.cedefop.europa.eu/en/news/italy-approval-national-implementation-plan-vet>
- Cedefop. (2025). *Vocational education and training policy briefs 2024 – Italy: National VET policy developments 2023-24*. Publications Office of the European Union.
- Cook, S., Watson, D., & Webb, R. (2024). Performance evaluation in teaching: Dissecting student evaluations in higher education. *Studies in Educational Evaluation*, 81, Article 101342. <https://doi.org/10.1016/j.stueduc.2024.101342>
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77(1), 113–143. <https://doi.org/10.3102/003465430298563>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.
- Drue, C. R., & Bifulco, C. A. (2025). Beyond the numbers: How directors and chairs interpret student feedback to equitably evaluate teaching. *Journal of Academic Ethics*, 23(3), 1499–1521. <https://doi.org/10.1007/s10805-025-09611-5>
- Dyer, K. D., & Donnelly-Hermosillo, D. (2024). Student ratings of instruction: Updating measures to reflect recent scholarship. *Research in Higher Education*, 65(7), 1587–1613. <https://doi.org/10.1007/s11162-024-09804-8>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–358. <https://doi.org/10.1037/h0061470>
- Gilbert, R. O., & Gilbert, D. R. (2025). Student evaluations of teaching do not reflect student learning: An observational study. *BMC Medical Education*, 25, Article 313. <https://doi.org/10.1186/s12909-025-06896-3>
- González-Fernández, R., Ruiz-Cabezas, A., Medina Domínguez, M. C., Subía-Álava, A. B., & Delgado Salazar, J. L. (2024). Teachers' teaching and professional competences assessment. *Evaluation and Program Planning*, 103, Article 102396. <https://doi.org/10.1016/j.evalprogplan.2023.102396>
- Heffernan, T. (2022). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 47(1), 144–154. <https://doi.org/10.1080/02602938.2021.1888075>
- Herning Declaration. (2025). *Herning Declaration on attractive and inclusive Vocational Education and Training (VET) for increased competitiveness and quality jobs 2026-2030*. Danish Presidency of the

- Hippe, R., Kyriakopoulou, P., & Psifidou, I. (2022). *Teachers and trainers in a changing world: Building up competences for inclusive, green and digitalised VET* (Cedefop Research Paper No. 86). Publications Office of the European Union.
- Holdsworth, S., Sandri, O., Hayes, J., & Maslen, S. (2025). Scenario-based capability assessment in professional learning contexts. *Journal of Adult and Continuing Education*, 31(3), 600–627. <https://doi.org/10.1177/14779714241297069>
- Isoré, M. (2009). *Teacher evaluation: Current practices in OECD countries and a literature review* (OECD Education Working Papers No. 23). OECD Publishing. <https://doi.org/10.1787/223283631428>
- Kim, F., Williams, L. A., Johnston, E. L., & Fan, Y. (2024). Bias intervention messaging in student evaluations of teaching: The role of gendered perceptions of bias. *Heliyon*, 10(17), Article e37140. <https://doi.org/10.1016/j.heliyon.2024.e37140>
- Le Boterf, G. (2008). *Repenser la compétence: pour dépasser les idées reçues, 15 propositions* (2e éd.). Éditions d'Organisation.
- Levinsson, H., Nilsson, A., Mårtensson, K., & Persson, S. D. (2024). Course design as a stronger predictor of student evaluation of quality and student engagement than teacher ratings. *Higher Education*, 88(5), 1997–2013. <https://doi.org/10.1007/s10734-024-01197-y>
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist*, 52(11), 1187–1197. <https://doi.org/10.1037/0003-066X.52.11.1187>
- Meijer, K., Baartman, L., Vermeulen, M., & de Bruijn, E. (2025). Mapping teachers' perceived importance of assessment competence: A quantitative exploration in the context of Dutch Higher Vocational Education. *Empirical Research in Vocational Education and Training*, 17, Article 10. <https://doi.org/10.1186/s40461-025-00185-5>
- Nussbaum, M. C. (2011). *Creating capabilities: The human development approach*. Harvard University Press.
- O'Donovan, R. (2024). Missing the forest for the trees: Investigating factors influencing student evaluations of teaching. *Assessment & Evaluation in Higher Education*, 49(4), 453–470. <https://doi.org/10.1080/02602938.2023.2266862>
- Pastré, P., Mayen, P., & Vergnaud, G. (2006). La didactique professionnelle. *Revue Française de Pédagogie*, 154, 145–198. <https://doi.org/10.4000/rfp.157>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE.
- Psifidou, I., Papazoglou, M., & Pouliakas, K. (2026). *VET teachers at a turning point: Pilot evidence from Cedefop's European Vocational Teacher Survey* (Cedefop Working Paper No. 29). Publications Office of the European Union. <https://doi.org/10.2801/1254726>
- Quansah, F., Cobbinah, A., Asamoah-Gyimah, K., & Hagan, J. E., Jr. (2024). Validity of student evaluation of teaching in higher education: A systematic review. *Frontiers in Education*, 9, Article 1329734. <https://doi.org/10.3389/feduc.2024.1329734>
- Röhl, S., Bijlsma, H., & Schwichow, M. (2025). Can feedback from students to teachers improve different dimensions of teaching quality in primary and secondary education? A hierarchical meta-analysis.

- Royal, K. D. (2017). A guide for making valid interpretations of student evaluation of teaching (SET) results. *Journal of Veterinary Medical Education*, 44(2), 316–322. <https://doi.org/10.3138/jvme.1215-201R>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Sen, A. (1999). *Development as freedom*. Oxford University Press.
- Søndergaard, M. K., Degn, L., & Lassesen, B. (2024). Thinking or talking about teaching? Student evaluation as an occasion for dialogue or reflection on teaching. *Tertiary Education and Management*, 30, 147–163. <https://doi.org/10.1007/s11233-024-09140-7>
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>
- Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1>
- Uttl, B., White, C. A., & Wong Gonzalez, D. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. <https://doi.org/10.1016/j.stueduc.2016.08.007>
- Vehovar, V., & Štrlekar, L. (2024). When to conduct student evaluation of teaching surveys: Before or after the final examination? *Assessment & Evaluation in Higher Education*, 49(6), 767–780. <https://doi.org/10.1080/02602938.2023.2298771>
- Woodzicka, J. A., Greer, L., Murdock, K. K., Johnson, D. R., Locy, T., & Goldsmith, A. H. (2025). One piece of the puzzle: Developing an empirically informed open-ended student evaluation of teaching. *Frontiers in Education*, 10, Article 1670426. <https://doi.org/10.3389/feduc.2025.1670426>