Evaluating Practicum Performance in ELT: Multi-Stakeholder Perspectives and Implications for Teacher Development

Ali Merç (ID)

Foreign Languages Education, Faculty of Education, Anadolu University, Turkey.

Abstract

Teaching practicum and its evaluation in language teacher education programs is a complex and multi-faceted phenomenon. This study investigated multiple perspectives on performance evaluation during the teaching practicum in an English Language Teacher Training Program by examining perspectives of university supervisors, cooperating teachers, peers, and the student teachers themselves. Utilizing a 45-item observation checklist and a combination of quantitative and qualitative analyses, findings revealed significant variance in evaluations across groups, with cooperating teachers providing the highest ratings. Interviews and group discussions highlight that while the checklist aids structured assessment, modifications are needed to accommodate diverse teaching contexts. Peer assessment and self-assessment were valued for fostering professional growth, though concerns about objectivity arose when linked to grading. The study suggests refining assessment tools, promoting reflective practices, and training evaluators to improve practicum experiences and support future teachers' development.

Keywords: Teaching Practicum; English Language Teaching (ELT); Teacher Development; Performance Evaluation.

1. Introduction

The teaching practicum component of teacher education programs is a dynamic and intricate process involving a range of participants including student teachers, cooperating teachers, university supervisors, and students. Assessing the performance of teacher candidates during practicum presents a significant challenge, given the diverse range of factors and considerations involved in this experience. The primary purpose of this study is to examine the diverse perspectives on performance evaluation within the teaching practicum, as viewed by various stakeholders: student teachers, their peers, mentor teachers, and university supervisors. Specifically, this research aims to identify the extent to which these groups perceive and value

performance evaluation differently. Additionally, the study seeks to explore student teachers' attitudes toward the observation and evaluation of teaching practices, as well as their views on self-assessment and peer assessment. By investigating these perspectives, the study intends to provide a comprehensive understanding of the evaluative practices in the practicum context and highlight areas for potential improvement in teacher education programs.

The literature identified several approaches to practicum assessment, each highlighting distinct aspects of teaching competencies. Subedi (2009) emphasized a comprehensive approach, including elements such as planning, preparation, classroom presentation, classroom management, communication skills, and evaluative techniques. The use of documentation portfolios was another established method, which allows for a reflective and longitudinal view of student teachers' development, as discussed by researchers like Goodman et al. (1989), Ryan and Kuhs (1993), and others (Shannon, 1994; Naizer, 1997; Gelinas, 1998; Rakow, 1999; Meeus et al., 2009). Additionally, Al-Mutawa and Al-Dabbous (1997) proposed an evaluation framework focusing on personal qualities, linguistic knowledge, interpersonal skills, and planning and implementation abilities. Similarly, Brooker et al. (1998) underscored the importance of planning and preparation, communication, teaching for learning, classroom management, student assessment, and professionalism in practicum assessment. Chen et al. (2011) introduced the Assessing Quality Teaching Rubrics (AQTR), which assesses task design, task presentation, classroom management, and response to students as core competencies. Good and Weaver's (2003) Professional Education Personnel Evaluation (PEPE) approach emphasized preparation, instructional organization, student performance assessment, classroom management, fostering a positive learning environment, communication, and leadership skills. Lastly, Darling-Hammond et al. (2013) proposed a holistic model encompassing planning, instruction, assessment, reflection, and academic language as essential components of effective practicum assessment. This overview of practicum assessment approaches highlights the multifaceted nature of teaching evaluations, suggesting that a blend of these components may provide a more comprehensive evaluation framework for student teachers. Within a study in Turkish context, Merç (2015) highlighted that while effective planning and preparation are vital for enhancing student teachers' practicum experiences, the evaluation process often faces challenges such as inconsistent qualifications of cooperating teachers, the undervaluation of peer feedback, and the limitations of report writing as an assessment tool.

All in all, this study addresses a critical gap in the literature by integrating and comparing diverse stakeholder perspectives on practicum performance evaluation, an area often neglected in existing research by answering the following research question:

- How do university supervisors, cooperating teachers, peers, and the student teachers themselves evaluate the performance of pre-service EFL teachers within the teaching practicum component of a language teacher education program?

2. Methodology

2.1. Setting and Participants

The study was conducted within the English Language Teacher Training Program at Anadolu University Faculty of Education, Türkiye. The participants were 30 fourth-year students, who were required to complete a 14-week practicum at a state secondary school within a *Teaching Practice* course. The practicum involved a commitment of six class hours per week, during which student teachers were expected to conduct a minimum of six teaching sessions. In this practicum setting, three student teachers were assigned to work under the guidance of a cooperating teacher. The university supervisor, who also served as the primary researcher, visited the practicum school twice throughout the duration of the course to observe and evaluate the performance of the student teachers, ensuring adherence to pedagogical standards and providing constructive feedback.

2.2. Data Collection

For the purpose of assessing the performance of teacher candidates during their practicum, an observation checklist was used. The checklist consisted of 45 items, organized into five distinct categories. *Preparation* (four items) focused on the extent to which the teacher candidate was adequately prepared for the lesson, including aspects such as lesson planning and organization. *Presentation* (ten items) assessed the candidate's ability to deliver content clearly, engage students, and effectively use instructional materials. *Execution* (thirteen items) evaluated the actual implementation of the lesson, including the ability to manage classroom dynamics, implement teaching strategies, and adapt to students' needs. *Personal characteristics* (eight items) examined the teacher candidate's professional demeanor, confidence, and communication skills. Finally, *Teacher-student interaction* (ten items) focused on the candidate's ability to foster a positive and productive relationship with students, promoting an environment conducive to learning. The checklist utilized a Likert-type scale for ratings: 1 = *Unsatisfactory*, 2 = *Average*, 3 = *Above average*, and 4 = *Excellent*.

The procedure for the research involved the completion of the observation checklist for each teaching performance. The checklist was filled out by each stakeholder. While the university supervisor and cooperating teachers evaluated the student teachers' performance based on their professional expertise and observations of instructional practices, student teachers self-assessed their own performance, reflecting on their lesson delivery and classroom management. Peer student teachers, who observed each other's lesson, offered assessments based on collaborative observations.

Following each teaching performance, group discussions as feedback sessions were held to provide constructive evaluations and highlight areas for further development. Finally, semi-

structured interviews were carried out with the participants to gather insights into their experiences with the observation process. The interviews included open-ended questions such as "*Do you think the checklist is appropriate for observing student teachers' lessons?*", "*How do you feel about observing and evaluating yourself?*", and "*How do you feel about observing and evaluating your peers?*"

2.3 Data Analysis

The data analysis for this study was conducted using both quantitative and qualitative methods to ensure a comprehensive examination of the teaching performance evaluations. For the quantitative analysis, *descriptive statistics* were first used to summarize the overall ratings across the five categories of the observation checklist. This allowed for an initial overview of the performance trends for each student teacher. Additionally, *One-Way Analysis of Variance (ANOVA) for repeated measures* was employed to examine whether there were significant differences in the evaluations provided by the four groups of participants. This analysis helped us determine whether the ratings varied across different observer perspectives and provided insight into the consistency or divergence of evaluations within each category. For the qualitative aspect of the analysis, a *data-driven methodology* was utilized. Semi-structured interview responses and feedback from group discussions were analyzed through thematic coding to identify recurring patterns and themes.

3. Findings

3.1. Quantitative Data

The descriptive analysis, presented in Table 1, highlights mean ratings across different assessment categories, as evaluated by four groups: The *University supervisor* (US), *Cooperating teachers* (CT), *Self-evaluations by student teachers* (ST), and *Peer student teachers* (PT). Each group assessed the five categories in the checklist: preparation, presentation, execution, personal characteristics, teacher-student interaction.

Table 1. Descriptive Analysis

	US	CT	ST	PT
Preparation	3,32	3,68	3,47	3,57
Presentation	3,01	3,65	3,48	3,61
Execution	3,17	3,61	3,39	3,49
Personal Characteristics	3,18	3,75	3,45	3,55
Teacher/Student Interaction	3,28	3,70	3,57	3,64
Overall	3,17	3,66	3,47	3,57

As Table 1 shows, *Preparation* received the highest average score from *CTs* (3.68), while *USs* rated it at 3.32. *Presentation* was rated highest by *CTs* (3.65) and lowest by *USs* (3.01). *Execution* was given a different rating across groups, with *CTs* scoring it at 3.61, while *USs* rated it at 3.17. *Personal Characteristics* showed a high score from *CTs* (3.75) but a lower rating from *USs* (3.18). *Teacher-Student Interaction* scored highest with *CTs* (3.70), with all groups showing relatively high agreement except USs (3.28). For the *Overall* category, *CTs* rated highest (3.66), while *STs* scored it slightly lower (3.47), and USs rated the lowest (3.17). These results suggest that CTs generally provided the highest ratings across most categories, particularly in *Preparation* and *Presentation*, while USs the lowest ones in all categories.

A one-way ANOVA analysis was conducted to examine the statistical significance of differences between groups in each assessment category, revealing that all differences were statistically significant at the *0.05 level*. Specifically, significant results were found for *Overall* ($F_{(3,87)}=11.311$, p=.000, *Preparation* ($F_{(3,87)}=3.542$, p=.018, *Presentation* ($F_{(3,87)}=17.515$, p=.000, *Execution* ($F_{(3,87)}=6.513$, p=.001, *Personal Characteristics* ($F_{(3,87)}=12.357$, p=.000), and *Teacher-Student Interaction* ($F_{(3,87)}=7.084$, p=.000). These findings indicate significant variance in ratings among the evaluator groups across all categories, with the most substantial differences observed in *Presentation, Personal Characteristics,* and *Overall* scores.

Post hoc tests were conducted to further explore these differences, identifying specific pairs where differences were significant at the 0.05 level, with some exceptions. In the *Overall* category, no significant difference was observed between *CTs* and *PTs*, or between *STs* and *PTs*. For *Preparation*, non-significant differences were noted between *US* and *STs*, *US* and *PTs*, *CTs* and *PTs*, and *STs* and *PTs*. Similarly, in the *Presentation* category, no significant differences were found between *CTs* and *STs*, *CTs* and *PTs*, and *STs* and *PTs*. For *Execution*, non-significant differences were observed between *US* and *ST*, *CT* and *PT*, and *ST* and *PT*. In *Personal Characteristics*, only *STs* and *PTs* showed non-significant differences, while in *Teacher-Student Interaction*, no significant differences were found between *CTs* and *STs*, *CTs* and *PTs*, and *STs* and *PTs*. These results underscore a level of consistency in certain ratings between specific evaluator groups, particularly between self and peer evaluations, which generally exhibited fewer significant differences compared to ratings from the university supervisor or cooperating teachers

3.2. Qualitative Data

The qualitative data collected through interviews and group discussions revealed several insights into the use of the observation checklist and its impact on the practicum experience.

Effectiveness of the Checklist for Observation and Evaluation: Overall, participants found the checklist to be a useful tool for observing and evaluating student teachers' teaching performance. One participant noted that the checklist provided a clear structure for assessing

the essential components of a lesson, stating, "*The checklist gives us an idea about the components and specific features of a lesson. Each item is like a guideline to check.*" Another participant reflected on how the checklist helped track their progress throughout the practicum, stating, "*I am even able to see my progress throughout the teaching practice, from the first lesson I delivered towards the end.*" This suggests that the checklist was not only a valuable source of immediate feedback but also contributed to long-term professional growth and self-reflection.

Need for Modifications to the Checklist: While the checklist was deemed helpful, some participants felt that it required modifications to better accommodate the diverse nature of teaching. One concern raised was that the checklist did not fully capture the variations in lesson structure. For example, a participant pointed out that certain types of lessons, such as those focused on language games or grammar presentations, might not align well with the existing checklist categories: "*What if all my lesson is allocated to playing a language game? No presentation, more classroom management issues. Or, what if I present a new grammar point for the whole 40 minutes? No group work, no pair work…*" This suggests that the checklist might benefit from greater specification regarding different teaching contexts, such as language skills (e.g., grammar, vocabulary, writing) or lesson types.

Importance of Self-Assessment for Awareness: Participants emphasized the importance of self-assessment in fostering greater awareness of their teaching practices. One participant shared that reviewing the checklist immediately after the lesson allowed them to reflect on their teaching, noting, "*When I look at the checklist just after I finish my lesson, I'm able to recall what happened there, and it's a good tool for reflection.*" Self-assessment not only helped participants identify areas they had planned but failed to achieve but also facilitated personal insights into their teaching performance. For instance, one participant mentioned learning the significance of knowing students' names, as highlighted by the checklist item "*Teacher knows students' names,*" reflecting on the importance of addressing students by name rather than using generic references like "*Yes, you, please.*"

Value of Peer Assessment: Peer assessment was viewed as a valuable tool for professional development, although some participants expressed initial discomfort with the process. One participant remarked, "*Yeah, first, it was difficult to evaluate my friend, but I know that this is for her own development as a teacher. So, I was not very generous.*" Despite this initial hesitation, participants acknowledged the importance of peer feedback in supporting mutual growth. However, concerns were raised regarding the use of peer assessment for grading purposes. As one participant noted, "*If it were for her grade to pass or fail, I wouldn't be, maybe, that much objective.*" This suggests that while peer assessment is valuable for development, it may require careful consideration when tied to formal grading criteria.

4. Conclusions

Based on the findings of this study, several suggestions and implications can be made to enhance the observation and evaluation process during the teaching practicum, as well as to improve the overall practicum experience for student teachers.

First, it is essential to develop and use valid and reliable checklists for the observation and evaluation of student teachers' performance. These checklists should be comprehensive and aligned with the key teaching competencies, ensuring that they accurately capture the multifaceted nature of teaching. Regularly updating and refining these tools will enhance their effectiveness in guiding evaluations. Second, holding group discussions after each teaching performance, involving university supervisors, cooperating teachers, and student teachers, can be an effective way to reflect on the teaching experience. These discussions provide an opportunity for constructive feedback and foster a collaborative learning environment, promoting professional growth for all involved. Third, establishing clear norms or benchmarks for each evaluation criterion can help ensure consistency and objectivity in the assessment process. By identifying what constitutes satisfactory, above-average, and excellent performance for each category, evaluators will be better equipped to provide targeted and meaningful feedback. Fourth, it is crucial to teach student teachers the importance of self-assessment as a tool for personal and professional development. Encouraging reflective practices can help student teachers identify their strengths and areas for improvement, leading to more intentional growth as educators. Fifth, teaching student teachers the value of peer-assessment can foster a sense of accountability and mutual support within the learning community. Peer feedback can offer alternative perspectives on teaching practices and can be particularly useful in enhancing collaborative learning environments. Sixth, providing training for cooperating teachers on how to effectively observe and evaluate student teachers' teaching can improve the quality of feedback provided during the practicum. A well-trained cooperating teacher is better equipped to give constructive, specific, and actionable feedback that supports the development of the student teacher. Seventh, shifting the focus from predominantly quantitative evaluation to a more qualitative approach will provide richer, more nuanced insights into a student teacher's performance. Qualitative feedback allows for the capture of complex aspects of teaching, such as student engagement, classroom atmosphere, and the ability to adapt to dynamic classroom situations. Finally, creating a less stressful teaching practicum experience will benefit both student teachers and evaluators. Reducing unnecessary pressure can foster a more supportive and productive learning environment, where student teachers feel empowered to experiment with new teaching strategies and reflect openly on their growth.

The above-mentioned suggestions, if implemented, have the potential to enhance the effectiveness of the practicum process, support the professional development of future educators, and ultimately contribute to the improvement of teaching practices in the classroom.

Acknowledgements

References

Al-Mutawa, N., & Al-Dabbous, J. (1997). Evaluation of student-teachers' performance in TEFL practicum at Kuwait University. *Journal of the Educational Research Center, 6*(11), 27–48.

Brooker, R., Muller, R., Mylonas, A., & Hansford, B. (1998). Improving the assessment of practice teaching: A criteria and standards framework. *Assessment & Evaluation in Higher Education, 23*(1), 5–24. https://doi.org/10.1080/0260293980230101

Chen, W., Hendricks, K., & Archibald, K. (2011). Assessing pre-service teachers' quality teaching practices. *Educational Research and Evaluation: An International Journal on Theory and Practice, 17*(1), 13–32. https://doi.org/10.1080/13803611.2011.578008

Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability, 25*(3), 179–204. https://doi.org/10.1007/s11092-013-9163-0

Good, J. M., & Weaver, A. (2003). Evaluation of preservice teachers' internships: A model to encourage career continuity and program reform. *Journal of Personnel Evaluation in Education, 17*(3), 263–275. https://doi.org/10.1007/s11092-005-2983-9

Goodman, K., Goodman, Y., & Hood, W. (1989). *The whole language evaluation handbook*. Heinemann.

Gelinas, A. M. (1998). *Issues of reliability and validity in using portfolio assessment to measure foreign language teacher performance* (Unpubl. doctoral diss.). The Ohio State University.

Meeus, W., Van Petegem, P., & Engels, N. (2009). Validity and reliability of portfolio assessment in pre-service teacher education. *Assessment & Evaluation in Higher Education, 34*(4), 401–413. https://doi.org/10.1080/02602930802062659

Merç, A. (2015). Assessing the performance in EFL teaching practicum: Student teachers' views. *International Journal of Higher Education, 4*(2), 44–56. https://doi.org/10.5430/ijhe.v4n2p44

Naizer, G. L. (1997). Validity and reliability issues of performance-portfolio assessment. *Action in Teacher Education, 18*(4), 1–9. https://doi.org/10.1080/01626620.1997.10463359

Rakow, S. J. (1999). Involving classroom teachers in the assessment of preservice intern portfolios. *Action in Teacher Education, 21*(1), 108–115. https://doi.org/10.1080/01626620.1999.10462951

Ryan, J. M., & Kuhs, T. M. (1993). Assessment of preservice teachers and the use of portfolios. *Theory into Practice, 32*(2), 75–81. https://doi.org/10.1080/00405849309543578

Shannon, D. M. (1994, July 10–15). *An evaluation approach for the development of preservice teachers* [Conference presentation]. Annual Meeting of CREATE (Center for Research on Educational Accountability and Teacher Evaluation) National Evaluation Institute, Gatlinburg, TN, United States.

Subedi, D. (2009). Assessing English language teaching student teachers on teaching practice. Journal of NELTA, 14(1), 138–145Bloom, B. S. (1956). *Taxonomy of Educational Objectives*. David McKay.