



Received: 12.11.2016
Received in revised form: 25.12.2016
Accepted: 07.01.2017

Horasan Dođan, S. & Cephe, P. T. (2017). A multi-perspective evaluation of teaching skills of ELT student teachers. *International Online Journal of Education and Teaching (IOJET)*, 4(1). 87-104.
<http://iojet.org/index.php/IOJET/article/view/161/152>

A MULTI-PERSPECTIVE EVALUATION OF TEACHING SKILLS OF ELT STUDENT TEACHERS

Seçil Horasan Dođan

Gazi University

secilhorasan@gazi.edu.tr

Pařa Tevfik Cephe

Gazi University

pcephe@gazi.edu.tr

Seçil Horasan Dođan is an English language instructor. She has currently been pursuing her PhD in ELT. She has had presentations at conferences and papers published. She is interested in teacher education, drama, and multi-culturalism.

Assoc. Prof. Dr. Pařa Tevfik Cephe is a faculty member in the ELT Department at Gazi University. His research interests include teacher education and TESOL methodology. He has presented papers at conferences and has published nationally and internationally.

Copyright by Informascope. Material published and so copyrighted may not be published elsewhere without the written permission of IOJET.

A MULTI-PERSPECTIVE EVALUATION OF TEACHING SKILLS OF ELT STUDENT TEACHERS

Seçil Horasan Doğan,

secilhorasan@gazi.edu.tr

Paşa Tefvik Cephe

pcephe@gazi.edu.tr

Abstract

Reflective practices in teacher education programs can play a critical role in enhancing the teaching quality. However, since student teachers' self-evaluations are not adequate on their own, they need to receive as much feedback as possible from other sources. In this study, a multi-perspective evaluation was provided to contribute to the teaching skills of student teachers. To this end, 15 ELT student teachers took part in an extra-curricular project to teach 1 lesson in a language classroom. Each lesson was recorded and evaluated first by themselves, then by the language learners in this classroom, and finally by 3 trainers based on the recordings using Teacher Evaluation Form. In addition, discussion sessions with each student teacher were held to share the multi-perspective evaluation and detect its contributions. The results of the quantitative and qualitative data analysis revealed that student teachers' self-evaluations were significantly different, yet they were lower than others unlike what was hypothesized. In addition, this reflective process contributed to their reflective skills, teaching skills, and self-awareness for their professional development.

Keywords: teaching skills, reflection, self-evaluation, video-based evaluation, a multi-perspective observation.

1. Introduction

In language learning, learners' achievement is highly dependent on teaching quality. Such practices as active engagement, learner-centered teaching, and constructivist learning all enable well-trained teachers to yield positive influence on language learning. However, the problem rises at the point where student teachers are challenged to apply these practices in the classroom, namely, to transfer the theoretical concepts into pedagogical implementations, which necessitates student teachers to be involved in more practices and to receive more feedback on their teaching practices. To solve the problem, educating reflective practitioners, as Schön (1987) proposed, has become the goal of teacher education programs to improve teaching quality and to set goals for professional development. In line with this requirement, constructivist frameworks have been incorporated in these programs by substantial number of teacher educators (Richardson, 1997; Walsh, 2003).

Educating reflective practitioners can be ensured through self-evaluation which can be made possible through reflections (Liou, 2001; Walsh, 2003; Ross & Bruce, 2007). Since reflection helps teachers recognize their limitations and discover perspectives and new choices, teachers' self-evaluation based on their teaching practices is highly important. Such reflections are believed to fill in the mentioned gap between theory and practice (Cephe,

2009). Since other means of reflection can be biased due to self-report effect, video-based self-evaluation works better for a sound reflection for being objective, effective, and efficient by enabling student teachers to analyze their own performances (Lee & Wu, 2006). That is, video-based evaluations prove to be more reliable than self-report evaluations. However, limited attention has been given to the use of videotaping as a reflective tool in teaching practice (Song & Catapano, 2008). In fact, the use of videos in teaching practice at teacher education programs is highly recommended for 3 main reasons: enabling visualization, facilitating reflection, and improving performances (Colasante, 2011). In addition, videotaping allows to observe the details missed in live observation in class; enables other observers to evaluate the same lesson; can be saved and watched as desired regardless of time and place; and provides an invaluable tool for reflection in professional development (Lee & Wu, 2006).

Self-evaluation alone, however, does not provide sufficient development to enhance teaching quality. As Colasante (2011) summarized, 3 components of critical reflection include peer discussion, teacher guidance and feedback, and linking theory to practice. In other words, as much feedback as possible should be provided to student teachers by their peers, course teachers, supervisors, or other sources about their teaching practice since such experience gives them the opportunity to apply theoretical knowledge to real practice (Lee & Wu, 2006). Even the evaluations of learners will contribute to the preparation of student teachers. Despite the value of the contributions and feedback from different sources, due to time constraints or lack of reflective practices, desired feedback from different perspectives cannot always be provided. Thus, there appeared a need to collect and compare evaluations from multiple perspectives about the teaching skills of student teachers so as to better contribute to their professional development. Since student teachers may be inclined to reflect subjectively on their own performances and evaluate themselves high in self-evaluation forms, such a comparison of 3 perspectives can enable them make more realistic and in-depth reflections about their teaching practices. In other words, because a multi-perspective evaluation is more reliable in providing evaluations from different aspects, it can contribute to student teachers who need to be trained to make objective reflections about themselves to improve their teaching skills. Thus, this study dwells on a multi-perspective evaluation for the purposes of providing multifaceted feedback, improving their teaching skills and awareness about them, and improving reflective skills.

The research was not within the teacher education program in order to eliminate any stress of being assessed by those with gate-keeper roles. Rather, the teaching performances of student teachers were observed in an intermediate level language classroom at tertiary level. Conceptualizing on reflective practices defined by Schön (1987) and constructivist views of Vygotsky (1978), this study aims to provide insights to teaching practices and professional development of student teachers through a comparison of 3 sources of evaluations: student teachers themselves, learners, and trainers. The study questions the following:

- Are there any significant differences among the evaluations of these 3 raters?
- How do a) student teachers evaluate their own teaching practice; b) language learners evaluate each student teacher; c) teacher trainers evaluate each student teacher?
- How does this multi-perspective evaluation contribute to the teaching skills of student teachers?

It is expected that there will be a difference in favor of student teachers who tend to evaluate themselves higher than other raters. The evaluations of learners and trainers are

expected to be parallel, as of the outsider eyes. Finally, multiple evaluations will contribute to the professional development of student teachers in terms of gaining awareness about teaching skills and reflective skills.

It can be stated as a limitation that case studies are limited to a group of participants making it difficult for generalization, and to a restricted time span of observation. In addition, item 1b in the instrument is assumed as the knowledge of learners and learning in this context.

2. Literature Review

2.1. Student teachers' self-evaluation

Teacher self-evaluation refers to 'a process in which teachers make judgments about the adequacy and effectiveness of their own knowledge, performance, beliefs, and effects for the purpose of self-improvement' (Airasian & Gullickson, 2005, 2). Self-evaluation has an essential role in teachers' professional development because it provides a major means for teachers to become aware of their own practices through reflections on experiences. As Schön (1987) denoted, learning from reflection distinguishes expert teachers from others. Reflection in this sense is the core of constructivist views for teachers in the process of constructing knowledge and beliefs based on cumulative experiences. He identified reflection-in-action related to the decision-making mechanisms during teaching, and reflection-on-action for any sort of evaluation after teaching. This study dwells more on reflection-on-action for student teachers' self-evaluation after watching their own teaching.

Reflective teaching entails teacher self-observation and self-evaluation in a cyclical way; that is to say, teachers manage to see what they cannot see. Reflection paves the way to fill in the gap between the actual teaching and the desired one. Accordingly, teachers who monitor their own teaching and evaluate themselves can easily make teaching decisions in designing their lessons. As a result, teacher reflectivity acts as a developmental process. Various means employed to this end trigger reflection and lead to self-discovery. Although reporting about the self can lead to bias, MacBeath (2003) advocated that teachers' self-evaluation is the most valuable and reliable source of information of what happens in the classroom. Critical reflection fosters awareness and understanding (Liou, 2001). Teacher self-evaluation does not only promote awareness; it also diverts teachers' attention on the ways to improve their practices for better. It fosters professional development and collaboration among colleagues. Their evaluations can then be compared to external evaluations. In terms of professional growth, Ross and Bruce (2007) underlined the role of teacher self-assessment as a constructivist tool in affecting excellence in teaching, enabling teachers to determine goals for development, enhancing communication with peers, and promoting external exchange of practices.

There are studies that touched upon teacher self-evaluation for varying motives. One is Walsh (2003) who highlighted reflective processes to increase the interactional awareness of second language teachers so as to redirect their interest to interactive decision making. Using guided self-discovery on conversation analysis, he interpreted the findings with teachers' self-efficacy beliefs. Similarly, Ross and Bruce (2007) argued that teachers' self-evaluations contribute to their self-efficacy beliefs. They also incorporated it with peer-coaching, observations, and feedback on teaching. Bullard (1998) emphasized teacher self-evaluation through the means of teaching portfolios, action research, journals, and the like. Lee and Wu (2006) advocated utilizing videos for student teachers' self-evaluation in web-based

computer-mediated communication based on the results that revealed effective improvement of teaching experience. Examining the role of reflective feedback, Eröz-Tuğa (2012) videotaped student teachers in their teaching practicum. She designed reflective feedback sessions on videotaped lessons and collected self-evaluation reports of student teachers on their teaching experiences. The results indicated positive effects of reflective feedback on teaching performances. She also claimed that the process lowered their anxiety, boosted self-confidence, and contributed to awareness. Liou (2001) investigated the reflective practice of student teachers on their own observation reports without using videos, and found no significant development in six weeks, which may have something to do with the developmental feature of video-based reflections.

2.2. Learners' Evaluation of Teaching

For all intents and purposes, the most important stakeholder in educational settings is the learners. Ur (1996) viewed learners as best sources for that purpose in that they can provide a holistic evaluation based on many lessons. As Kurtoğlu Eken (1999, 240) stated 'we have a lot to learn from our learners.' Thus, it can be inferred that the most worthwhile feedback on the effectiveness of teaching can be obtained from learners who are teachers' obvious critics. Historically dating back to Remmers in the 1920s (Wachtel, 1998), literature literally abounds in studies on language learners' evaluation of instruction. However, most are on the overall assessment of a course, particularly at universities serving to the purposes of course feedback, tenure decisions, and promotion (Zabaleta, 2007). Nevertheless, language learners' evaluations of teachers based on specific observation are not many. Particularly those on student teachers, to our best knowledge, are none as they have fewer chances of teaching in real classrooms, with those who have being at younger groups who may not be eligible to evaluate teachers.

It has to be acknowledged that language learners' feedback on teachers and the effectiveness of instruction has many benefits (Wachtel, 1998; Ballantyne, Borthwick, & Packer, 2000). For one, their feedback can enhance the quality of instruction as they play a diagnostic role. Ratings of language learners are the most commonly applied measure of teaching effectiveness. Teachers can receive immediate feedback from learners to gauge the effectiveness of different techniques they apply. Also learners are more powerful in today's schools for decision-making. They feel valued when their opinions are consulted. Furthermore, a combination of learners', teachers' feedback, and observers' feedback can yield more reliable consequences for decision-making. Cooperation with learners in this way helps establish a better relationship and a classroom culture. Such collaboration is also useful for the observers to compare their feedback with those of others. Marsh (1987) reported that learner ratings are multidimensional, useful for all stakeholders, and reliable for not being contaminated by bias. Overall, needs should be seen from the eyes of learners by schools to reach a higher quality (Ballantyne, et al., 2000).

2.3. Observers' Evaluation of Teaching

Observation, particularly collaborative observation, is very important in professional development. An observer could be any stakeholder. Marsh and Roche (1997) offered that language learners, former language learners, teachers themselves, colleagues, trained observers, or administrators can evaluate teaching effectiveness. Mentors, supervisors, peers, experienced or expert teachers can also be the observers. These people take on the observer role to evaluate one's teaching performance rather than to make judgments. They interpret actions within their contexts rather than simply watching and reporting. To this end, the best

means to contribute to professional development can be videotaped reflective teaching practices (Lee & Wu, 2006; Song & Catapano, 2008; Colasante, 2011; Eröz-Tuğa, 2012). As Schön (1987, 303) underlined, ‘a dialogue of reciprocal reflection-in-action between coach and student’ is the key to train reflective practitioners. In this study, the observers are trainers.

2.4. A Multi-Perspective Evaluation of Teaching

Literature does not abound in multi-perspective evaluations of teaching practices. One of the studies that dwelled on learners’ feedback in addition to other-party feedback was by Richards (1998), a pioneer in three-way observation studies, who combined teachers’, peers’, and learners’ feedback. In three way observations, in addition to professional development practices in which teachers became pairs and observed their lessons for peer-evaluation, learners’ feedback was also gathered. On the other hand, assuming teachers and learners as two must-parties in class by nature, Shortland (2004) described peers as the third-party observation.

In one such study, Kurtoğlu Eken (1999) aimed to find out the role of learners’ feedback in improving teaching and learning quality. In each lesson to be observed, one learner took on the learner observer role to take notes and share them later on. She found that learners’ feedback and suggestions are quite supportive in exploratory practice by giving some directions to teachers. Learner observers, though they provide simpler feedback, can even systematically contribute to teachers. Although with different groups, Ozogul, Oline, and Sullivan (2008) designed 3 groups of student teachers who were involved in teacher-evaluation, self-evaluation, and peer-evaluation groups to investigate the improvements in planning lessons. They expectedly found that those in the teacher-evaluation group showed the most development. Yet, the importance of other sources cannot be underestimated.

In brief, a study on differences among student teachers’ self-evaluation, the evaluation of learners they teach, and that of observers, say peers, expert teachers, teacher educators, or trainers, can provide insights for professional development. In this way, student teachers can gain awareness on their teaching skills, notice their strengths and weaknesses, modify their practices accordingly, and shape their beliefs through reflective experiences.

3. Methods

3.1. Research design

In this study, a case study methodology was adapted to explore the evaluations of 3 groups of participants on each of their teaching performances. A particular instance of a more general situation, a case study provides a rich description of a case, its analysis, perceptions of the participants, relevant events, and a clear understanding of ideas (Cohen, Manion, & Marrison, 2007). This study is an exploratory case study based on observational data. The teaching performance of each student teacher was observed and evaluated by themselves, language learners, and trainers, one being the researcher. Cohen et al. (2007) stated that a description of events is blended with their analysis in case studies in which the researcher is also involved to seek the perceptions of individuals or groups on events. A mixed methodology, which enables to obtain rounded and reliable data (Cohen et al., 2007), was used to gather quantitative data from rubric evaluation and qualitative data from reflection questions and final discussions.

3.2. Participants

There were 3 participant groups who were selected through convenience sampling due to their availability and accessibility at the time (Cohen, et al., 2007). The first was 15 student teachers who were all female seniors at English Language Teaching (ELT) at a public university. The second group was 15 intermediate level language learners at tertiary level aged mostly 18. They were studying in the same classroom at the time. Due to absenteeism of some learners during evaluations, the number of learners evaluating each teacher varied from 7 to 15. The third group was 3 ELT instructors with varying amount of experience from 7 years to 11 years at the preparatory language program at a university. They all had ELT background, were pursuing postgraduate degrees in ELT or Curriculum, and joined teacher training courses.

3.3. Instruments

An evaluation form called the Teaching Evaluation Form was adapted from Danielson's Framework for Teaching (2007), the most commonly used and adapted framework (Kimball & Milanowski, 2009). Danielson based this framework of teaching practice on constructivist approaches and designed it to identify the performance levels that teachers are expected to show, and illustrated them with classroom-based examples. The framework sets a standardized evaluation of teaching practice (Kimball & Milanowski, 2009). Thus, it has been highly acknowledged by teacher training programs. Because of ready-made behaviorally referenced scales of the framework that cover the complexity of teaching, the teacher evaluation systems, particularly in the United States, commonly used the framework or its variations depending on their purposes (Milanowski, 2011). One of the studies that adapted the framework for reflective professional development purposes was conducted by Song and Catapano (2008) who designed a twenty-four-item survey of 4-point scale selecting from the elements of the first three components to evaluate videotaped lessons of 8 in-service teachers. Three external reviewers rated the same lessons and found 0.972 of reliability in Cronbach's Alpha computation. They concluded that ratings were almost identical.

Danielson's Framework for Teaching (2007) is composed of 4 domains and 22 components. Domain 1 is Planning and Preparation including 6 components in which knowledge of content, pedagogy, learners, materials, and objectives are mainly addressed. Domain 2 is the Classroom Environment including 5 components in which classroom procedures such as smooth transitions, encouragement, physical space and safety, and respect are addressed. Domain 3 is Instruction including 5 components in which clear communication, discussion, active engagement of learners, monitoring, flexibility, critical thinking, and decision-making are addressed. Domain 4 is Professionalism including 6 components in which responsibilities of teachers for professional development such as their cooperation with colleagues are addressed. All domains include five or six components and each component has several elements. The components are described by detailed indicators, possible examples, and critical attributes.

Of all domains, Domain 2 and 3 are directly relevant to classroom observation and Domain 1 can be traced through lesson planning and the evidence in classroom. However, since Domain 4 pertains to professional development that takes place outside the classroom in longer terms and in relation to other stakeholders, it was left out in the adaptation. As a result, Danielson's first 3 domains including sixteen components with described performance levels as Unsatisfactory, Basic, Proficient, and Distinguished graded as 1 to 4 respectively

were adapted as are for the purposes of this study. Having 4 scales as in Danielson's is more preferred in rubrics so as to avoid the central tendency effect (Popham, 2005).

In addition to these rubric questions, 3 open-ended questions were added for the participants to reflect and provide more feedback about the teaching (Appendix). These questions were composed of reflective questions used in similar studies (see Kurtoğlu Eken, 1999; Song & Catapano, 2008). Thus, the instrument was composed of an observation scale as originally presented by Danielson's first 3 domains and a feedback tool with open-ended reflection questions. Therefore, it was eluded that the instrument did not guide or structure the responses of any raters in a particular way.

All participants evaluated the same lesson using the same tool. Although the framework is extremely commonly adapted, expert opinion was consulted for the last version and raters' evaluations were piloted on different samples of recordings first. As Song and Catapano (2008) suggested, practical reliability ensured by training the raters is more valuable than computing reliability. Thus, the rubric and form were sent to the participants in advance and they were also trained on how to use it referring to the indicators, examples, and critical attributes that Danielson described. For reliability, the Cronbach's Alpha was computed as 88.1% for student teachers, 91.9% among learners, and 91.3% among trainers. As Cohen, et al. (2007) stated, there may be bias in self-reporting, which can be eliminated through triangulation or other means because it is critical to avoid a threat to external validation. In addition to the rubric forms, a final reflective discussion session was held with each student teacher to discuss the comparative evaluations to track their contributions and were audio-recorded.

3.4. Procedure and Data Analysis

In the planning phase, the teaching dates were determined with the student teachers who were invited to teach one lesson as guest teachers. Therefore, the second question in Domain 1 about knowing learners might have been low. Thus, it was assumed as the knowledge of learners and learning in general and the ways teachers sought ways to get to know about the learners. All student teachers taught in the same class. The lesson objectives and materials were shared with them in advance, and then they sent their lesson plans to the trainers. Before the lesson, language learners were informed about the guest teachers. In the training phase of the raters, the language learners were trained in their native language to ensure their understanding and they were also assisted through more examples and translations during evaluations.

In the data collection phase, each lesson was video-recorded. After the lessons, the videos were first shared with the student teachers to evaluate their own performance on the rubric. Secondly, each learner was given the rubric to evaluate the guest teacher while watching the lesson in the classroom. It is important that such experimental studies are conducted on out of the curriculum and testing procedures of the classroom so as not to bear grading anxiety and bias so that any sort of reciprocity effect (Clayson, Frost, & Sheffet, 2006), a situation in which learners give high evaluations to teachers who gave them high grades and vice versa, can be eliminated. Finally, 3 trainers watched the videos and evaluated the student teachers on blind rating. A few disparities were negotiated and disagreements were resolved. All raters were encouraged to write as much comments as they could in the last part of the instrument. Gathering the scores, the researcher trainer discussed the comparative evaluations with the student teachers to encourage them to reflect more on their teaching and to interpret the

evaluations about themselves so as to see how beneficial this process was for their professional development and what they gained at the end.

In the data analysis phase, all participants were coded such as ST1...ST15 for student teachers, LL1...LL15 for language learners, and TT1...TT3 for teacher trainers. After coding, the quantitative data were entered to SPSS. First, student teachers' self-evaluations were directly keyed to the database. Secondly, learners' evaluation scores for each student teacher were grouped. The extreme ends in learners' data – the highest and the lowest scores – were discarded to compute the trimmed mean. In robust statistics, trimmed mean removes outliers for more objective and robust analysis so that more consistent and accurate distributions can be found (Larson-Hall, 2010). The trimmed mean scores of learners were also keyed to the database. Finally, for the consistency of trainers, their scores were entered to SPSS for computing the rater agreement on Kendall's W test for reliability concerns. This test explores the agreement among at least 3 non-continuous variables (Hatch & Lazaraton, 1991). In this study, 3 independent trainers rated the same performances based on the same instrument they were trained about. In Kendall's W test for the consistency of 3 trainers' ratings, $KW = 0.832$ was found. As 0 means no agreement while 1 means complete agreement according to $0 \leq W \leq 1$, the result was accepted to provide sufficient agreement. It shows that groups have a similar way of thinking in evaluating performances, deriving most importantly from the training before the study. The importance of training the participants in using the instrument plays a critical role to increase reliability in this sense because raters' understanding the tool is much more critical than the numbers computing reliability (Song & Catapano, 2008). Consequently, the mean scores of the 3 were computed to be used as group scores.

The data from all 3 groups were analyzed on Shapiro-Wilk Normality Test for normality since it is a powerful goodness-of-fit test in small sized-samples (Larson-Hall, 2010). In Shapiro-Wilk test, null hypothesis is not rejected when p is greater than .05 (Larson-Hall, 2010). The analysis in Table 1 shows that a normal distribution was found.

Table 1. *Shapiro-Wilk Test of Normality*

		Tests of Normality		
	Raters	Shapiro-Wilk		
		Statistic	df	Sig.
Mean scores	Trainers	.928	15	.258
	Student teachers	.917	15	.174
	Learners	.969	15	.842

Following these phases, Levene's Test was administered for homogeneity and one-way ANOVA was used to determine any differences and the multiple comparison test was applied for further analysis. Finally, the constant comparison method was used for qualitative data analysis. That is, the data were continuously re-examined to compare them within groups and the theoretical assumptions (Cohen et al., 2007). It was revealed in general that language learners' and trainers' evaluations match to 80%. The audio-recordings of final discussions were transcribed and analyzed with the same method.

4. Results

The first research question investigated the differences in a multi-perspective analysis to contribute better to the teaching skills of student teachers. Following the reliability analysis,

further analysis was conducted through one-way ANOVA for testing whether any statistical differences exist among three groups (Larson-Hall, 2010). It showed that .05 was accepted as significant, and there was a difference among participants as in Table 2. In addition, Levene test, in which significance level of more than .05 indicates equal variances, was conducted for homogeneity of variance (Larson-Hall, 2010). Table 3 shows variable homogeneity was found.

Table 2. Results from the one-way ANOVA test

ANOVA					
Mean scores	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	268.876	2	134.438	3.18	.05
Within Groups	1776.791	42	42.305		
Total	2045.667	44			

Table 3: Results from the Levene Test of Homogeneity

Test of Homogeneity of Variances				
Mean scores	df1	df2	Sig.	
Levene Statistic	.360	2	.700	

Having more power in finding differences especially among three means, Least Significant Difference (LSD) Test was applied (Table 4) as the most powerful test in post-hoc multiple comparison tests (Larson-Hall, 2010). The results showed a significant difference in favor of student teachers. The trainers' and learners' scores were close to one another and higher than those of student teachers while the student teachers' scores for themselves were relatively lower in contrast to what was hypothesized.

Table 4. Results from the LSD Test for multiple comparisons

Multiple Comparisons						
Dependent Variable: mean scores						
(I) Rater	(J) Rater	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
trainers	student teachers	5.73333*	2.37500	.020	.9404	10.5263
	learners	1.37190	2.37500	.567	-3.4210	6.1648
LS student teachers	trainers	-5.73333*	2.37500	.020	-10.5263	-.9404
	learners	-4.36143	2.37500	.073	-9.1544	.4315
learners	trainers	-1.37190	2.37500	.567	-6.1648	3.4210
	student teachers	4.36143	2.37500	.073	-.4315	9.1544

*. The mean difference is significant at the 0.05 level.

The second research question investigated the perceptions of three rater groups in evaluating the same teaching practice of each student teacher. First of all, student teachers' self-evaluations all showed that they all enjoyed the lesson and the experience, though they also highlighted some of their weaknesses to be improved. They all stated that such real classroom experiences should be more frequent as they can well contribute to their teaching skills prior to the profession. In addition, they all reported the usefulness and effectiveness of receiving feedback from different perspectives. One point about the qualitative and quantitative data they provided was that their perceptions in reflection and scores in the forms were mostly parallel in %73 of them. One of the student teachers (ST14), for instance, scored herself and was scored by others as a medium-achiever. She commented *'The students were eager to join the lesson. However, I should improve my teaching skills'*; one learner (LL3) commented *'It was an enjoyable lesson, but she was shouting while speaking'*; and one trainer (TT2) commented *'She had several language mistakes, but a high energy to activate all learners.'* This showed three parallel perspectives on ST14. However, they were not always parallel to those of the trainers and learners. For example, one teacher (ST4) was criticized by the trainers and learners, but she stated *'It was an enjoyable and effective lesson and a great practice before starting the profession'*. Another (ST13) who was highly appreciated underscored herself stating that *'I am a bit weak to handle unexpected situations.'*

Learners, on the other hand, were quite frank in their evaluations. When they liked the lesson, the activities, and the teacher, they commented positively as *'LL4: She was the best teacher of all. LL8: She spoke fluently. LL9: We thought she was experienced. LL2: She was confident and enjoyable. LL16: She will be an active and beloved teacher.'* for one teacher (ST10) or negatively for another (ST1) as *'LL5: Her looks were harsh. LL7: She was talking as if she would beat us at any moment. LL6: She made us nervous. LL1: She was energetic but cold.'* The evaluations of learners were not so shallow as to state that it was a good lesson. In contrast, they used the items in the rubric while referring to the teaching skills so that they constituted another critical rater. To illustrate, for ST3, LL8 said, *"She was too excited as understood from her body language and she could not use her voice effectively"* while LL16 said, *"Her movements and teaching were like memorized, not natural."*

As for the trainers, the comments were expectedly more comprehensive and to the point. For ST9, for example, TT3 commented *'She had a good rapport with learners, but she gives the feeling that she may be challenged under unexpected situations.'* For ST13, TT1 stated *'Her use of authentic materials was not only pursuant but also effective for permanent vocabulary learning. However, she did not ask learners any questions for comprehension check. Rather, she directly required some production. It created the feeling that she focused too much on what she had to do mechanically, and underestimated how much students learned.'* Their evaluations and learners' were in line. Just like learners, TT1 and TT3 commented for ST3 as *"Her excitement was obvious, yet she was aware and trying to handle it. Her smooth transition made the lesson connected. However, somehow she covered the lesson literally, which seemed so unnatural."*

When ratings from three different perspectives were examined for each teacher, the most interesting result was that ST10 whose self-evaluation was the lowest of all was evaluated as the most successful by trainers and the second most successful by learners. Similarly, ST9 with the second lowest self-evaluation was rated relatively high by trainers and learners. In contrast, ST6 who scored herself in the second highest rank was evaluated low both by the trainers and learners and ST4 in the fourth rank among self-evaluations was the lowest by

trainers and the second lowest by learners. These four cases were the most strikingly different ones. However, the consistency between the trainers and learners was still apparent.

The third question investigated the possible contributions of the multi-perspective evaluation on student teachers. To this end, the final discussions with each student teacher about the comparative evaluations of multiple perspectives were productive in that they had the opportunity to compare their perceptions to those of other sources that can be more objective. Their reactions and responses showed that they were more excited to hear the evaluations of language learners. This was the first time they had feedback from their learners. They stated that they discovered their weaknesses better and they sought ways to cope with them. For instance, ST1 stated *“I knew that my eye shots are extra hard as my professors at my teacher training program always tell me; however, I did not know that it affected the learners that much. I try hard to diminish it, but what else can I do? Could you please advise me something about this?”* She was upset to hear the evaluations of learners, yet was more willing than ever to change her looks. She was noted that she was one of those who raised awareness the most, particularly in Domain 2 and 3. She was highly appreciated in planning and adaptations, yet she became aware of the importance of creating a friendly atmosphere and having a good rapport with learners thanks to the evaluations of learners in particular. She was suggested mirror practice for a while.

On the other hand, ST10, one of the extreme cases, had considered herself weak in this particular teaching practice; thus, she was shocked and happy to learn the high evaluations of the learners and trainers. She said *“I underscored myself because I could not manage my time well and I thought they could not learn the subject. However, I see that they felt comfortable with me, learned the target words, and had thought that I was experienced. It means I could have created this atmosphere, but I was not aware. I guess I was afraid or not confident enough. Now, I wish I can teach more in this class.”* Her gain was more on self-awareness. She realized the importance of self-assessment, monitoring, responsiveness, and flexibility in class.

Another extreme case, ST4, became aware of her weaknesses not immediately after watching the video, but after the discussions. She stated *“I used LI in class to explain better. While watching the video, I thought it was OK; but now while negotiating with you on the video, your evaluations, and learners’ evaluations...I do not know. I think I need more practices. I realized the flaws in my plan, instructions and body language on the video; however, I can only now understand that it caused a distance between me and the learners. They told me “a cold teacher”, but I am always friendly in my normal life.”* She noticed that her activity selection was one thing that affected her lesson and the implementation of these activities was another. Even if she had selected coherent activities based on lesson objectives, they might not have been influential unless she established a comfortable environment and communication with learners and modified her instruction in a more interactive and responsive way. Thus, she was concluded to come to realizations in all domains of the evaluation.

The other student teachers also indicated that this teaching experience was fruitful to receive evaluations from different perspectives, to discover their strengths and weaknesses, and to negotiate ways to improve their teaching. Most of them were good at Domain 1 as seen in lesson planning, but not in Domain 2 (a friendly atmosphere, encouragement, classroom management, physical organization, and smooth transitions) and Domain 3 (clear explanations, thought-provoking questions, active participation, monitoring, and responsiveness in unexpected situations). Thus, they discovered they needed to improve

teaching skills on classroom environment and instruction. They became aware of what they did in class and how their practices were perceived by learners. In addition, they appreciated the way it was conducted outside their coursework so that both student teachers and learners eluded the anxiety of being assessed.

5. Discussion

Drawing from the results, one important finding was that the hypothesis regarding the first research question did not come true. Due to the self-report effect, student teachers were expected to score themselves higher. However, interestingly enough the quantitative analysis showed that whilst the trainers and learners scored student teachers higher, they scored themselves quite low. This is not congruent with Ross and Bruce (2007) who stated that it was not very likely for teachers that underrated themselves to change because of the depressing impact of negative self-evaluation. However, it can be explained in that the external threat of self-report had been eliminated by asking the teachers to watch themselves on recordings and make the evaluations accordingly. Therefore, having seen themselves objectively, student teachers provided honest reflections. This strengthens the power of video-based reflections and evaluations in improving teaching skills of student teachers (see Colasante, 2011).

Regarding the next hypothesis, the results about how each rater group evaluated the same student teacher showed that as outsider evaluators, the trainers and learners provided parallel evaluations to the student teachers. As expected, outsider eyes were compatible while student teachers' reflections were harsher due to objective evaluation, yet were much lower than expected. Therefore, it can be argued that through video-based evaluations self-report bias can be avoided so that self-assessment could become objective and effective thanks to video-based reflections, which is compatible with what Lee and Wu (2006) argued.

In addition to the consistency among rater groups, contributions of each can be discussed. To being with student teachers themselves, except for the four extreme cases, the result that 73% of student teachers scored themselves in the form in parallel to the way they evaluated themselves in the reflective questions can be valued high considering that they took limited part in practicum or any other real teaching contexts although they were used to microteaching. This is an indication of high reliability. Secondly, trainers' comments were more guiding, enlightening, and informative; thus they contributed a lot to student teachers in raising awareness for various aspects of teaching practices. Finally, learners revealed meaningful evaluations that would help student teachers figure out their powers and flaws supposedly thanks to the training they received about using the instrument. In addition, some of their evaluations were outspoken allegedly because they took their evaluation task seriously when they were informed that this project served to the professional development of the student teachers. Learners' contribution to the process was also proven to be extremely supportive by Kurtoğlu Eken (1999).

In respect to the third question, one of the most important findings arose in the analysis of the final discussions that basically revealed the contributions of this process to student teachers. Student teachers had the chance to receive a comparative evaluation from multi-perspectives. They were also themselves constituted one of the perspectives through their own reflections. Therefore, they had the opportunity to learn the ways to make more objective reflections. In addition, they became better aware of their strengths and weaknesses so that they can take a step to develop themselves. They discovered their own mistakes in the videos. However, most importantly they received feedback from various perspectives so that

even if they had missed a point in the video or did not understand the value of a point, they could have realized it at the end. As Schön (1987) described in reflection-on-action, student teachers can be involved in a conscious, non-spontaneous evaluation of the whole picture once the lesson is over to make more long-term decisions on different aspects of their teaching practices.

Furthermore, they particularly looked forward to hearing the evaluations of learners since it was the first time they were evaluated by learners for many reasons. First, they are always evaluated by their teacher educators or peers in microteaching; thus, their comments do not make as much effect as learners'. Second, they have limited experiences in practicum to allocate time for learner evaluations. Third, they usually teach to young learners in practicum who are not mature enough for objective evaluations. As a result, student teachers all stated that learners' comments were more influential on them to take actions to modify their teaching. This is a significant outcome to lead to some research on to learner evaluation.

Finally, the last hypothesis verified that this processes contributed to student teachers who were pleased to seek ways to compensate their weaknesses and fortify their strengths. They were more willing to take part in real teaching contexts to improve their teaching skills through experience because they believed that learning by doing or experiencing is extremely valuable. Some of them could not show or develop certain teaching skills, yet they certainly indicated and also stated that they had at least gained awareness not only on teaching skills but also on reflective skills. The results showed that the least development and contributions were observed in Domain 1 about planning and preparation probably due to numerous practices of lesson planning and microteaching in their teacher education. However, the most development and awareness were observed in Domain 2 (classroom environment) and Domain 3 (instruction) most probably due to the opportunities of practicing in real classroom contexts and reflecting not only on their own performances after watching videos but also on the multi-perspective evaluations compared in the final discussions. This shows the significance of more teaching practices (Seferoğlu, 2006) and reflective feedback (Eröz-Tuğa, 2012).

Therefore, this study suggests a developmental process for student teachers by allowing them to take part in more practice by teaching outside their teacher education program, enabling them reflect on their video-recorded teaching practices, engaging them in multiple evaluations, and encouraging them to reflect on the whole process for self-discovery. Reflecting upon teaching and observing duality in actual teaching settings contribute a lot to student teachers for further professional development.

6. Conclusion

A comparison of multi-perspective evaluations was examined in this study. The results showed that despite the self-report effect, student teachers reported themselves lower than the trainers and learners did. It is argued that this is a consequence of video-based reflections that makes scoring more objective and effective. This finding is compatible with what Lee and Wu (2006), Song and Catapano (2008), and Colasante (2011) argued.

Regarding the reflective aspect of the study, it can be concluded that when the opportunity is provided to student teachers to be involved in self-evaluation, it is seen that they tend to discover certain things about themselves. This shows the importance of reflection for professional development once again as already asserted in literature (Eröz-Tuğa, 2012; Cephe, 2009; Schön, 1987) and a need for more and more teaching practice (Seferoğlu,

2006). Therefore, to enhance the reflective practices of student teachers, they should be encouraged to be involved in video-based practices, to keep reflective journals, and to take part in discussions with colleagues, peers, mentors, and even the learners they teach. Since it was found that student teachers discovered certain things not when they watched the videos alone, but when they did so with a trainer with whom they also negotiated the evaluations, it can be concluded that teachers need both emotional and intellectual support; thus, affective dimension should be addressed (Le Cornu & Ewing, 2008). In order to address their affective modes, more collaborative observations are needed. Thus, to improve teaching skills, collaborative observations and reflections are more beneficial. This is congruent with what Atay (2008) discussed with collaborative dialogue for professional development.

In respect to the multiple-perspective evaluations, it can be concluded that student teachers' self-evaluations were important for self-awareness and reflection; learners' evaluations were quite motivating and supportive; trainers' evaluations were extremely beneficial and eye-opening. One conclusion that can be drawn out of the findings from the final discussions on the comparative evaluations of three perspectives is that triangulation of evaluations makes feedback to student teachers more versatile, wealthy, supportive, and useful. A comparative evaluation demonstrates if there is a difference between the evaluations of themselves and those of others so that they can improve their reflective skills for objectivity and self-betterment. This is similar to what Schön (1987) discussed with reflection-on-action through which holistic, conscious, non-spontaneous evaluations as well as healthy long-term decisions can be made.

Self-evaluation and supervisory evaluations are rather prevalent in literature, but the learner dimension in the evaluations, which is not very common for pre-service teachers, was confirmed to be quite useful and necessary because student teachers, having almost no experience of teaching in real classroom contexts, became curious to hear learners' evaluations and motivated to make modifications on their teaching even if the feedback was positive or negative. What they remembered more and influenced them more at the end of the process was what learners commented about them. Therefore, as long as the learners are eligible and mature enough and are trained to evaluate student teachers, their evaluations are seen to be quite significant for the professional development of student teachers. This is in line with Kurtoğlu Eken (1999) who found that learners could contribute to teachers and it was really supportive.

Finally, the rating tool can be employed with the student teachers in different fields of teaching as well. As a limitation, however, it can be stated that the teaching performances of student teachers were evaluated based on their one simple lesson by three perspectives. Therefore, it would be harsh to make radical judgments about their teaching. A further study can be conducted on several subsequent observations. Furthermore, as the fourth perspective, peer-evaluation can be incorporated to future studies just like the one of the components in three-way observations suggested by Richards (1998) and Shortland (2004).

Acknowledgements

We would like to express our appreciation to Ferudun Sezgin for his support in the statistical analysis.

References

- Airasian, P. W., & Gullickson, A. R. (2005). Teacher self-evaluation. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (pp. 186-211). California: Corwin Press.
- Atay, D. (2008). Teacher research for professional development. *ELT Journal*, 62(2), 139-147.
- Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching: Identifying and addressing academic staff development needs. *Assessment & Evaluation in Higher Education*, 25(3), 221-236.
- Bullard, B. (1998). *Teacher self-evaluation*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association Cambridge: Harvard University Press.
- Cephe, P. T. (2009). An analysis of the impact of reflective teaching on the beliefs of teacher trainees. *Education and Science*, 34(152): 182-191.
- Clayson, D., Frost, T., & Sheffet, M. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education*, 5(1), 52-65.
- Cohen, Louis, Manion, L., & K. Morrison. (2007). *Research Methods in Education (Sixth Edition)*. London & New York: Routledge.
- Colasante, M. (2011). Using video annotation to reflect on and evaluate physical education student teaching practice. *Australasian Journal of Educational Technology*, 27(1), 66-88.
- Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria: ASCD.
- Eröz-Tuğa, B. (2012). Reflective feedback sessions using video recordings. *ELT Journal*, 67(2), 175-183.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, Heinle & Heinle.
- Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.
- Kurtoğlu Eken, D. (1999). Through the eyes of the learner: Learner observations of teaching and learning. *ELT Journal*, 53(4), 240-248.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. NY, Routledge.
- Le Cornu, R., & Ewing, R. (2008). Reconceptualising professional experiences in student teacher education...reconstructing the past to embrace the future. *Teaching and Teacher Education*, 24(7), 1799-1812.

- Lee, G. C., & Wu, C. C. (2006). Enhancing the teaching experience of pre-service teachers through the use of videos in web-based computer-mediated communication (CMC). *Innovations in Education and Teaching International*, 43(4), 369-380.
- Liou, H. C. (2001). Reflective practice in a student teacher education program for high school English teachers in Taiwan, ROC. *System*, 29(2), 197-208.
- MacBeath, J. (2003). Teacher self-evaluation. In *International Handbook of Educational Research in the Asia-Pacific Region*, pp. 767-780. Volume 11 of the series *Springer International Handbooks of Education*: Springer Netherlands.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253-388.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187.
- Milanowski, A. T. (2011). Validity Research on Teacher Evaluation Systems Based on the Framework for Teaching. *ERIC*, 520519. Retrieved from <http://eric.ed.gov/?id=ED520519>
- Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development*, 56(2), 181-201.
- Popham, W. J. (2005). *Classroom assessment: what teachers need to know*. Boston: Pearson/Allyn and Bacon.
- Richards, J. C. (1998). *Beyond training*. Cambridge: Cambridge University.
- Richardson, V. (Ed.). (1997). *Constructivist teacher education: Building a world of new understandings*. Washington, DC: The Falmer.
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146-159.
- Schön, D. A. (1987) *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey Bass.
- Seferoğlu, G. (2006). Teacher candidates' reflections on some components of a pre-service English teacher education programme in Turkey. *Journal of Education for Teaching*, 32, 369- 378.
- Shortland, S. (2004). Peer observation: A tool for staff development or compliance? *Journal of Further and Higher Education*, 28(2), 219-228.
- Song, K. H., & Catapano, S. (2008). Reflective professional development for urban teachers through videotaping and guided assessment. *Journal of In-Service Education*, 34(1), 75-95.

- Ur, P. (1996). *A course in language teaching: Practice and theory*. Melbourne: Cambridge University.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-212.
- Walsh, S. (2003). Developing interactional awareness in the second language classroom through teacher self-evaluation. *Language Awareness*, 12(2), 124-142.
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12(1), 55-76.

Appendix: Teacher Evaluation Form Adapted from Danielson's Framework for Teaching**Teacher Evaluation Form***Is this:*

Your name (evaluator):

Teacher's self-evaluation:

Name of the teacher observed:

Student evaluation of teacher:

Aim of the lesson:

Observer's (i.e. peer, trainer) evaluation:

Instruction: Please watch the videotaped lesson and evaluate the teaching according to the scale:
 1=Unsatisfactory, 2=Basic, 3=Proficient, 4=Distinguished.

Domain 1: Planning and Preparation	1	2	3	4
1a: The teacher demonstrates knowledge of content and pedagogy				
1b: The teacher demonstrates knowledge of students				
1c: The teacher sets goals and being prepared				
1d: The teacher demonstrates knowledge of resources (materials and technology)				
1e: The teacher designs coherent instruction by selecting varied, appropriate activities				
1f: The teacher designs student assessments according to their performance				
<i>Total for Domain 1:</i>				
Domain 2: The Classroom Environment				
2a: The teacher creates an environment of respect and rapport through interaction and body language				
2b: The teacher establishes a culture for learning with pride and encouragement				
2c: The teacher manages classroom procedures from grouping to smooth transitions				
2d: The teacher manages student behavior with a respectful manner				
2e: The teacher organizing physical space to be safe and accessible				
<i>Total for Domain 2:</i>				
Domain 3: Instruction				
3a: The teacher communicates with students through clear instructions and explanations				
3b: The teacher uses thought-provoking questioning and discussion techniques				
3c: The teacher engages students in learning actively				
3d: The teacher uses assessment in instruction through monitoring and giving feedback				
3e: The teacher demonstrates flexibility, responsiveness, and adjustment to unexpected situations				
<i>Total for Domain 3:</i>				
Total				
Reflection				
1. How did you feel about the lesson/teaching?				
2. What were the strong and weak aspects of the lesson/teacher?				
3. Do you have any suggestions for this lesson?				