ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

Assessment of Teaching Performance: Mentors and Student Self-Rating

Kim Francis R. Rodriguez
Student Teaching Mentor, Cebu Normal University - Integrated Laboratory School

ABSTRACT

This paper examined the assessment by pre-service teachers (PSTs) and mentors on student teaching performance. Specifically, the study determined the mentors' rating on student teaching performance, the student teachers' self-evaluation, and the gap between the two ratings. PSTs and mentors from the elementary level of the Cebu Normal University-Integrated Laboratory School participated in the study. A performance appraisal sheet was utilized to gauge the teaching performance of the PSTs. Results indicate that student teachers and mentors offer varying perspectives in assessing teaching performance. PSTs rate themselves significantly higher than their mentors. There is a call for communicating to the assessor and the assessed a clear and specific set of standards against which student teaching performance will be assessed. This will not only address the gap but will also encourage an appraisal encompassing the intentions of both the PSTs and mentors. For the PSTs, an appraisal to promote learning, and for the mentors, one that checks conformity to pre-established standards.

Key Words: *assessment, student teaching performance, self-rating, mentor's rating*

INTRODUCTION

Practicum experiences play an indispensable part in the earliest phase of becoming a teacher. Pre-service teachers (PSTs) consider the practicum experience their litmus test before their initiation to the teaching front. It is not surprising therefore that during practicum, PSTs worldwide express a wide range of concerns one of which is on how their practice teaching is being assessed.

Assessment is a recognized valuable tool to promote learning (Assessment Reform Group, 2002; Black and William, 1998; & Shephard, 2000 as cited by Tillema, 2009). It checks the conformity to marked achievements from pre-established standards (Zuzowsky & Libman, 2002; & Heilbronn et al., as cited by Tillema, 2009). However, practicum assessment programs are revealed to be conceptually problematic (Brooker et al., 1998). Primarily, assessors assess a teaching performance based on their own intentions and what they consider relevant. Consequently, 2 angles need to be looked into: for the PSTs, a learning oriented appraisal, one in which support and guidance are taken into consideration; and for the mentors, an appraisal assessing performance improvement, one that is based on strict standards (Tillema, 2009).

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

A PST's rating is measured based on mentors' pre-determined criteria. Falchikov & Magin (1997), Guilford (1965), and Newstead & Dennis (1994 as cited by Falchikov and Goldfinch, 2000) express that mentors' marks alone pose a problem. Newstead and Dennis (1990, as cited by Falchikov and Goldfinch, 2000) contend that these marks are not fair due to different kinds of biases occurring during the assessment period. To address this concern, Wheeler and Knoop (1982) explored the feasibility of including self-assessment along with the mentors' assessment in determining a PST's rating. They found out that PSTs' self-evaluation were significantly higher than their mentors' rating. This disagreement between mentors' and the PSTs' self assessment is further validated by Tillema in 2009. This discrepancy between the two raters makes the ratings unreliable thus, not objective. Rothstein (1989) contends that objective procedure is one in which agreement among raters is at maximum.

Existing literature on assessment of student teaching performance focuses on degrees of agreement among multiple raters – self, mentor, and field supervisor (Tillema, 2009; and Wheeler & Knoop, 1982). Several studies indicate that different stakeholders hold a wide variety of perspectives on appraising PSTs during practice teaching (Atwater & Brett, 2005; Tillema & Smith, 2007; Wilson & Youngs, 2005; Zuzowsky & Libman, 2002 as cited by Tillema, 2009). Appreciation of a multirater assessment derives from the contention that no single source in the appraisal of teaching performance has ultimate legitimacy or warranty (Cochran, Smith, & Fries, 2002 as cited by Tillema, 2009). Others, such as studies of Falchikov & Magin (1997), Guilford (1965) and Newstead & Dennis (1994) (as cited in Falchikov and Goldfinch, 2000) focus on the validity and reliability of mentor markings on student teaching performance. On the other hand, Falchikov and Goldfinch (2000) compared peer and teacher marks. All these studies share a common goal, to investigate on an objective assessment of student teaching performance, one in which the congruence among raters (no matter how many they are) is at its highest.

Ideally, assessors in a multirater assessment should have an agreement in their standards they are assessing in teaching performance. In reality, this is not the case. Wheeler and Knoop (1982 as cited by Anderson et al., 1995) found out that ratings from mentors and self do not guarantee objectivity, as teachers tend to give low marks, and PSTs tend to overrate their performance. Furthermore, Tillema (2009) found out that there is considerable variation in purposes and intentions among assessors in a multirater assessment. This current study aims to find out if this same trend is still true in this modern era.

Study Objectives

This paper examined the assessment by PSTs and mentors on student teaching performance.

Specifically, the study determined the:

1. mentors' rating on student teaching performance by subject [Math, English, Science, Filipino, HEKASI (Heyograpiya, Kasaysayan, at Sibika), MAPEH (Music, Arts,

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

Physical Education, and Health), MTB-MLE (Mother Tongue-Based Multilingual Education)];
2. PSTs' self evaluation by subject (Math, English, Science, Filipino, HEKASI, MAPEH, MTB-MLE);&
3. gap in the performance rating between the mentors and PSTs;

METHODOLOGY

Study Design

This study utilized a quantitative method of inquiry. Specifically, it used a descriptive-comparative design as it determined the difference between the scores given by the mentors and the PSTs' self-rating on their demonstration teaching.

Participants

The participants of this study were the PSTs and the student teaching mentors (STMs) in the Integrated Laboratory School (ILS) in Cebu Normal University (CNU). Using maximum variation purposive sampling, the PST respondents were chosen from among the Bachelor of Elementary Education (BEEd) PSTs who are: (1) enrolled in the current semester; (2) assigned in the elementary level; & (3) in their in-campus exposure. The STMs of the chosen STs automatically became the STM respondents as they are the ones who would rate the PSTs in their teaching demonstrations.

Research Locale

The study was conducted in the Integrated Laboratory School of Cebu Normal University. The laboratory school is headed by a supervisor with a teaching force of 26 STMs. Most STMs are assigned as advisers to specific grade levels. These grade levels are assigned with PSTs who teach the pupils for one whole semester.

Research Instruments

In determining the self-rating and mentors' rating on PSTs' teaching demonstration, the appraisal sheet of the CNU-ILS was utilized. This appraisal sheet is composed of 4 components namely lesson planning, strategies of teaching, communication classroom management, and communication skills. Each of these components have certain indicators which are rated using Likert scale values of 2 – poor; 4 – moderately satisfactory; 6 – satisfactory; 8 – very satisfactory; and 10 – excellent.

Data Gathering Procedure

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

A letter seeking permission to have the PSTs and their STMs participate in the study was sent to the supervisor of the ILS. Once permission was obtained, PST participants were selected though maximum variation purposive sampling. The selected PSTs were asked to do a self-rating of every teaching demonstration he/she will have. These self-ratings were collected along with the ratings given by the STMs on the same teaching demonstrations. Using simple mean, the researcher determined the teaching performance of the PSTs as rated by both the STMs and by themselves. A two-sample t-test was used to determine whether there exists a significant difference between these two ratings.

RESULTS AND DISCUSSION

Assessment aims at providing informative feedback to help the PST gain insight into performance so that it is valuable to his or her professional growth (Boshuizen, Bromme, & Gruber, 2004). Traditionally, mentors hold the responsibility of appraising a student teaching performance based on pre-established standards. However, literature reveals that PST's self-rating do not always agree with their mentors' – an observation that dates back to the 80's (Wheeler & Knoop, 1982; Tillema, 2009). The findings of this study deal with (a) mentors' ratings on PST performance, (b) the PST's self rating, and the (c) gap between these two ratings.

Mentors' Rating on Student Teaching Performance

Mentors rate the PSTs' teaching performance along 4 components namely – lesson planning, strategies of teaching, classroom management, and communication skills. Table 1 presents the weighted mean of their ratings to the PSTs' teaching performance per subject.

Table 1
Mentors' rating on students teaching performance by subject

Subject	No. of Demos	Mean	SD	Description
Filipino	10	8.31	0.98	Very Satisfactory
Math	12	8.23	1.00	Very Satisfactory
MAPEH	11	8.10	1.27	Very Satisfactory
Values	7	8.12	1.31	Very Satisfactory
Science	11	8.01	1.22	Very Satisfactory
HEKASI	10	8.24	0.59	Very Satisfactory
English	8	7.08	2.07	Satisfactory
MTB-MLE	3	8.20	1.04	Very Satisfactory
HELE	5	9.43	0.29	Outstanding

Note. The ranges for the weighted mean are: 0.00-2.99 – Unsatisfactory; 3.00-4.69 – Fair; 4.70-7.99 – Satisfactory;
8.00-9.29 - Very Satisfactory; & 9.30-10.00 – Outstanding

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

Among the 9 subjects (with the number of teaching demonstrations ranging from 3 to 12 per subject), only one of them was rated 'Outstanding', 7 were rated 'Very satisfactory', and one was rated 'Satisfactory'. The data show that STMs in the different subjects vary in their ratings to the PSTs teaching in the subject. Naturally, this is because not all PSTs have the same teaching capabilities. Some of them may be good in planning the lesson but has difficulty managing the class, or good in managing the class yet lacks the necessary communication skills needed to effectively deliver the lesson. In the English subject for example, it can be noted that the SD is high compared to the rest of the subjects. This is because half of those demonstrations were rated 'Very Satisfactory' and above while the other half was rated 'Satisfactory' and below. For one, this might be due to the PSTs' lack of the necessary communication skills to effectively deliver the assigned subject matter this being a language subject. Moreover, this could also be due to the PSTs' inability to submit their lesson plans (LPs) early resulting to the LPs not getting approved, thus failing on the lesson plan component of the appraisal sheet.

Generally, the PSTs have a very satisfactory teaching performance as indicated by the mean scores of all teaching demonstrations in the 9 subject areas. The varying rating provided by the STMs mean that each mentor has a different set of perspectives, each one informed by its own intentions and what he or she considers as important. Tillema (2009) found out that "assessment is a process closely linked to assessors' intentions and the aspects the assessor considers relevant". Some STMs might put premium on the quality of the lesson plan presented before the demonstration, others on the communication skills, while others emphasize on classroom management. This indicates and validates a long-standing problem in appraising student teaching performance – the lack of articulation of the criteria for assessing practicum (Brooker, R., Muller, R., Mylonas, A., & Hansford, B., 1998). Clear and specific set of criteria and grading procedures are not made explicit to the PSTs, STMs and field supervisors. Tillema (2009) found out that in the appraisal of student teaching performance, there's always a lack of clarity of goals and transparency of procedures. She further indicated that the greatest discrepancy is in the competencies weighted as indicative of teaching performance. Wheeler and Knoop (1982) suggest the presence of a halo effect, that is, the lack of differentiation among distinct categories of competencies tested in student teaching.

Overall, the PSTs have a very satisfactory teaching performance in the 8 subject areas based on the standards of the STMs.

Student Teachers' Self-Evaluation by Subject

PSTs rate themselves in their teaching demonstrations along 9 identified subject areas. Table 2 presents the data from among 77 teaching demonstrations in these subjects.

Table 2
Student teacher self-rating on their performance by subject

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

Subject	No. of Demos	Mean	SD	Description
Filipino	10	8.93	0.38	Very Satisfactory
Math	12	8.78	0.52	Very Satisfactory
MAPEH	11	8.43	1.14	Very Satisfactory
Values	7	8.64	0.70	Very Satisfactory
Science	11	8.74	0.97	Very Satisfactory
HEKASI	10	8.89	0.66	Very Satisfactory
English	8	8.07	1.63	Very Satisfactory
MTB-MLE	3	8.35	0.29	Very Satisfactory
HELE	5	9.21	0.53	Very Satisfactory

Note. The ranges for the weighted mean are: 0.00-2.99 – Unsatisfactory; 3.00-4.69 – Fair; 4.70-7.99 – Satisfactory;
8.00-9.29 - Very Satisfactory; 9.30-10.00 – Outstanding

Generally, the PSTs rated themselves very satisfactory in their teaching performance in all subject areas. It is easily noticeable that all subjects were rated 'Very Satisfactory'. Looking at the mean and the standard deviation, we can see that the PSTs have a higher consensus as far as their self-rating is concerned. On average, they rate themselves high. This finding runs parallel with that of Wheeler and Knoop's (1982 as cited by Anderson et al., 1995) indicating that PSTs' self-evaluation were significantly higher than either academic or field supervisors' rating. This can be attributed to the fact that PSTs consider the practicum experience a paramount activity in their teacher education program thereby giving their best in every teaching demonstration they have. This indicates that the PSTs maximize their learning experience in the practicum thereby involving themselves in constructing judgments that inform their chosen field (Dochy, Segers, & Sluijsmans, 1999).

In a review conducted by Dochy et al. (1999) on self, peer, and co-assessment in higher education, he found out that research reports positive findings concerning the use of self-assessment in educational practice. Further, he posited that self-assessment leads to more reflection on one's own work, a higher standard of outcomes, responsibility for one's own learning, and increasing understanding of problem-solving. In the process of doing self-rating, the PSTs become aware of the criteria they are tested against in every teaching demonstration. This allows them ample time for preparations thus resulting to a more sound and well-prepared teaching demonstration.

Overall, the PSTs have a very satisfactory teaching performance in all subject areas based on their own ratings.

Gaps in the Performance Rating between and Mentors and Student Teachers

ISSN: 2278-4012, Volume:8, Issue:1, JANUARY 2019

PSTs' self-rating and STM's rating on the same teaching demonstration enriche the assessment process as it provides multiple-perspective viewpoints offered by the one doing the actual job and the other – a mere observer. Table 3 presents the mean difference in the ratings provided by the STMs and the PSTs themselves.

Table 3
Mean gap (difference) in the performance between mentors and student teachers

Rater	Mean	SD	T-Value	P-Value
Mentor	8.14	1.23	-3.15	0.002**
Student Teacher	8.65	0.89		

Note. ** - highly significant at $\alpha = 0.01$

PSTs significantly have higher rating for themselves than their STMs. There is a high likelihood that PSTs put higher self rating in their teaching performance. This runs parallel with the findings of Wheeler and Knoop (1982) in their study on self, teacher and faculty assessments of student teaching performance. They found out that PSTs overrate their performance while STMs underrate them. PSTs tend to be more lenient than their STMs and that STM ratings consistently evaluate PSTs according to some global judgment.

The results may also indicate a dissonance in the standards set and the appraisal focus in the actual practice lessons. A central question to this indication is whether STMs and PSTs employ a concerted and aligned assessment in learning to teach (Tillema, 2009). Assessment is a process closely linked to assessors' intentions and the aspects they consider relevant (Tillema, 2007). STMs assess PSTs' performance based on strict standards while PSTs seek a learning orientation in appraisal. Students ask for a supportive, guidance-oriented assessment rather than an appraisal based on strict standards. (Tillema, 2009). Addressing this query may align what the STMs are looking for in a teaching demonstration and what the PSTs should do in order to exhibit what is sought for. Further, it encourages PSTs to accept feedbacks from mentors easily and follow recommendations given by the latter. This informative assessment is believed to improve PSTs' motivation and self-esteem because it improves their learning (Falchikov, 2005).

CONCLUSION AND RECOMMENDATIONS

PSTs and STMs offer varying perspectives in assessing teaching performance. PSTs rate themselves significantly higher than their mentors. There is a call for communicating to the assessor and the assessed a clear and specific set of standards against which student teaching performance will be assessed. This will not only address the gap but will also encourage an appraisal encompassing the intentions of both the PSTs and the STMs. For the PSTs, an

appraisal to promote learning, and for the STMs, one that checks conformity to pre-established standards.

Recommendation for research on the inclusion of an intermediary factor in addressing the gap is also offered. While there's a considerable literature on the exploration of a mutlirater approach in appraising student teaching performance, none explores on taking into consideration the test scores of the pupils in reconciling the gap between STMs and PSTs' self-rating in determining the latter's final rating.

REFERENCES

Anderson, J. B., & Freiberg, H. J. (1995). Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly*. 22:1, 77-91

Boshuizen, H. P. A., Bromme, R., & Gruber, H. (2004). *Professional Learning: Gaps and transitions on the way from novice to expert.* Dordrecht, Netherlands: Kluwer Academic.

Brooker, R., Muller, R., Mylonas, A., & Hansford, B. (1998). Improving the assessment of practice teaching: A criteria and standards framework. *Assessment and Evaluation in Higher Education.* 23:1 5-20

Dochy, F., Segers M., & Sluijsmans, D. (1999): The use of self-, peer and coassessment in higher education: A review, *Studies in Higher Education*. 24:3, 331-350 doi:10.1080/0307507991233137993 5

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* London: Routledge Falmer.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*. 70:3 287-322

Guaglianone, C. L., Payne, M., Kinsey, G. W., & Chiero R. (2009). Teaching performance assessment: A comparative study of implementation and impact amongst California state university campuses. *Issues in Teacher Education.* 18:1 129-148

Hudson, P. (2010). Mentors report on their own mentoring practices. *Australian Journal of Teacher Education.* 35:7, 30-42.

Leijen, Ä., Slof, B., Malva, L., Hunt, P., Tartwijk, J., & Schaaf, M. (2017). Performance-based competency requirements for student teachers and how to assess them. *International Journal of Information and Education Technology.* 7:3, 190-194 doi:10.18178/ijiet.2017.7.3.864

Newstead, S., & Dennis, I. (1995). Blind marking and sex bias in student assessment. *Assessment and Evaluation in Higher Education.* 15, 132-1399

Rothstein, J. M. (1989). On defining subjective and objective measurements. *Physical Therapy.* 69:7, 577-579

Tillema, H. (2009). Assessment for learning to teach: appraisal of practice teaching lessons by mentors. *Journal of Teacher Education*. 60:155, doi:10.1177/0022487108330551

Tillema, H. H. & Smith, K. (2007). Protfolio assessment: In search of criteria. *Teaching and Teacher Education*, 23(4), 442-456.

Wheeler, A. E. & Knoop, H. R. (1982). Self, teacher and faculty assessments of student teaching performance. *The Journal of Educational Research.* 75:3, 178-181, doi:10.1080/00220671.1982.10885377