

Use of Survey Data to Evaluate Teaching: A Comparison of Self and Peer Evaluations of Teaching

Ugorji I. Ogbonnaya

Tshwane University of Technology, South Africa
E-mail: ogbonnayai@tut.ac.za

Doi:10.5901/mjss.2014.v5n9p370

Abstract

Survey is one of the most widely used methods of investigation into educational systems and practices. While the results from surveys might give insight into what goes on in the educational system, important questions arise: how reliable are the results of surveys in evaluating educational systems and practices? And can the results be used to take high-stakes decisions about teaching and learning? This study examines the reliability of surveys in evaluating teaching by comparing the degree of agreement between mathematics teachers' self-evaluation of their teaching and their peers' evaluation of it. A sample of 31 grade 11 mathematics teachers took part in the study. The result shows that teachers' self-evaluations of their teaching generally did not agree significantly with their peers' evaluations. This finding suggests caution in the use of self and peer evaluations in taking high-stakes decisions on teaching and learning.

Keywords: Mathematics teaching, peer evaluation, self-evaluation, evaluation of teaching, survey

1. Introduction

Instructional quality is an important factor that affects student learning (Kersting, Givvin, Thompson, Santagata & Stigler, 2012). Therefore, evaluating teachers' classroom instructional practices is an important aspect of evaluating the quality of the teaching that they provide for the students. It is also an important aspect of studying the implementation and effects of education policies. Often, to evaluate classroom instruction, large-scale studies rely on representative survey data to describe it (Desimone, Smith & Frisvold, 2010; Praetorius, Pauli, Reusser & Rakoczy, 2014). Responses to evaluation surveys of classroom instructional practices depend on the respondents' perceptions of what makes classroom instruction effective. This is because what makes teaching effective is a subject of much debate and dialogue (Ko, Sammons & Bakkum, 2013). Hence, respondents may differ in their opinions on the effectiveness of classroom instruction. Understanding that there might be response differences among individual respondents will help us understand the strengths and weaknesses of survey data we use to study classroom instructional practices (Desimone, Smith & Frisvold, 2010). This is found to be important now because the last few decades have seen a proliferation of surveys in education research studies; without reliable results no study can meaningfully contribute to knowledge and any decision based on such a study will be misguided.

2. Background

Evaluation of educational systems and practices and in particular teachers' instructional practices has been an age-old practice in education. An educational system is evaluated to obtain feedback on the functionality of the system in order to improve the system and for the purpose of accountability to the stakeholders (Patton, 1987). Evaluation of teachers' teaching is appraisal of the effectiveness of their pedagogical practices. It is used to: (1) provide formative feedback to teachers for improving the content and structure of their teaching; (2) measure their teaching effectiveness for promotion and retention purposes; (3) monitor the quality of instruction provided for students; (4) obtain information about instructional practices that support high standards and good practice, and (5) obtain data for research on teaching (Chen & Hoshower, 2003; Denson, Loveday & Dalton, 2010).

Self and peer evaluations are among the most commonly used forms of evaluation of teaching practices. Self-evaluation of teaching is a process whereby a teacher carries out self-appraisal of his/her teaching. It enables the teacher to judge his/her instructional practices in retrospect. Self-evaluation of one's instructional practices is an important source of evidence of one's teaching (Cranton, 2001), it demonstrates one's knowledge about one's perceived effectiveness in

the classroom and can present a picture of one's teaching that is unobtainable from any other source (Berk, 2005).

Peer evaluation of teaching is a process whereby a teacher's instructional practices are observed, reviewed and judged by the his or her peers. Praetorius, Pauli, Reusser and Rakoczy (2014) noted that observer rating, for example peer rating of instructional practices, is considered in some studies as the best option for evaluating teaching and is sometimes included as a component of instructional research. Peer evaluation of teaching as affirmed by French-Lazovik (1981) provides valuable evidence of the person's teaching from other teachers (colleagues) who are knowledgeable about the subject matter and the teaching contexts. Peer evaluation of a teacher's pedagogy involves observation of her or his classroom instructional practices by peers, appraisal of the teaching and learning materials he or she used and giving constructive commentaries on the teacher's teaching practices.

Often, because of the ease of data collection and the minimal cost of the processes, self and peer evaluations of teaching are used to appraise the effectiveness of instructional practices. However, there is continual debate on the trustworthiness of the result obtained through survey studies in teaching evaluation (Desimone, Smith & Frisvold, 2010). One concern about the use of self and peer evaluations is the possible biased estimate, amongst both groups, of teaching (Berk, 2005). Also, as observed by Praetorius, et al. (2014), the quality of observed lessons on which instructional evaluations are based may not be sufficiently indicative of the teachers' general teaching conduct. This concern becomes heightened when results from surveys are to be used in policy formation or in high-stakes decisions about teaching.

3. Research Question

This paper aims to compare the degree of agreement between mathematics teachers' self-evaluation of their teaching and their peers' evaluation of it. Hence, the research questions addressed are (1) is there any difference between teachers' self-evaluation of their instructional practices and their peers' evaluation of the teachers' instructional practices? And (2) to what extent do teachers' self-evaluations of their instructional practices correlate with their peers' evaluations of the teachers' instructional practices?

4. Method

4.1 Participants

The sample consisted of 60 high school mathematics teachers from one education district in a province in South Africa. The sample was a convenience one because the participants consisted of only the teachers who volunteered to take part in the larger study from which this paper is drawn. The participants were allowed to withdraw from participation at will; as a result, some of the teachers withdrew before the completion of the study. Hence, the sample after deletion of missing data was 33. Of the 33 teachers 11 taught their classes and hence undertook the self-evaluation of their teaching while 22 (2 for each teacher) observed the 11 teachers' teaching and carried out the peer evaluation.

4.2 Instruments

Two questionnaires (a self-evaluation questionnaire and a peer evaluation questionnaire) were used to collect data from the teachers and their peers.

4.2.1 Self-evaluation questionnaire

The instrument was a teacher self-report six point Likert scale (6 = strongly agree, 5 = agree, 4 = slightly agree, 3 = slightly disagree, 2 = disagree, 1 = strongly disagree) questionnaire developed to seek teachers' information on issues pertaining to their teaching, namely, their knowledge of the subject (TK), lesson preparation (LP), lesson organisation (LO), lesson presentation (LP_r) and student assessment (SA).

4.2.1.1 Development of the instrument

In developing the instruments, a framework of effective mathematics teaching was formulated as suggested by Berk (1979). Then, using the framework, the following steps recommended by Marsh & Hocevar (1991) were followed:

- Development of a large pool of items (from literature, existing instruments, interviews with students and

teachers),

- Piloting the instrument to receive feedback about the items, and
- Consideration of the psychometric qualities of the items while revisions are made.

The process is discussed in detail elsewhere (see Ogbonnaya & Mogari, 2011). However, items in the final instrument are briefly recorded in Table 1. The categorisation of the items is not dichotomous as some items were found to belong to more than one category in the development and validation process but were placed in the category in which they received the highest rating from the experts who validated the instrument and the statistical process that followed. For example, 'I used different resource materials' (LPr8) was also a lesson preparation variable but it was rated higher under lesson presentation than lesson preparation by the experts. The items are classified into five categories as shown in Table 1.

Table 1: Items in the teacher self-evaluation questionnaire

	<i>Item name</i>	<i>teacher self-evaluation item</i>
Knowledge of content	K1	I demonstrated sound knowledge of the curriculum
	K2	I demonstrated sound knowledge of students' prior conceptions
	K3	I clearly explained procedures for solving problems
	K4	I explained concepts in different ways to help my students understand
	K5	I used a variety of instructional strategies appropriate to the topic
	K6	I always related ideas to students' prior knowledge
Lesson preparation	LP1	I planned in advance to ensure lessons run smoothly
	LP2	I always attended class
	LP3	I was always punctual for class
Lesson presentation	LPr1	I reviewed lessons by connecting to previous classes
	LPr2	I previewed lessons by connecting to future classes
	LPr3	I created opportunity for students' participation in classroom problem solving
	LPr4	I carried the whole class along in my lesson delivery
	LPr5	I asked probing questions when a student answer was incomplete
	LPr6	I encouraged my students to ask questions
	LPr7	I supported lessons with useful class work
	LPr8	I used different resource materials
	LPr9	I brought lessons home by using examples in the students' context
	LPr10	I built my teaching on students' prior knowledge
	LPr11	I gave individual support to students when needed
	LPr12	I adjusted lessons when my students experienced difficulties in learning
	LPr13	I challenged my students to think through and solve problems, either by themselves or together as a group
	LPr14	I took extra steps to help all my students learn and achieve success in maths
	LPr15	I taught in a manner that took cognisance of differing learning abilities of students
Lesson organisation	LO1	I ensured that lessons on the topic were well organised
	LO2	I provided an inspirational learning environment
	LO3	I well articulated teacher tasks and student tasks
Student assessment	SA1	I asked questions to monitor students' progress
	SA2	I monitored and checked students homework always
	SA3	I gave feedback to students about their homework and assignment
	SA4	I appropriately integrated assessment in teaching
	SA5	I used assessment results to provide extra help to my students

4.2.2 Peer evaluation questionnaire

The peer evaluation questionnaire was used by the colleagues (peers) of the participating teachers to evaluate the teachers' instructional practices. It is similar to the self-evaluation questionnaire. To ensure consistency in the two instruments, the 32 items in the self-evaluation questionnaire were all retained in the peer evaluation questionnaire but the first person pronouns used in the self evaluation instrument were changed to third person pronouns in the peer evaluation instrument. For example, the item '*I explained concepts in different ways to help my students understand*' (K4) in the self-evaluation instrument was changed to '*the teacher explained concepts in different ways to help his/her students understand*' in the peer evaluation instrument.

4.2.3 Instruments' validity and reliability

Construct validity of the instruments was ascertained through judgment by experts. They rated the appropriateness of each item to the construct it was purported to measure, and the average rating of each item was computed and correlated with between-total scores across all items in the construct (Trochim, 2006). Only the items with correlation coefficients greater than or equal to 0.7 (Trochim, 2006) were selected for the instruments.

The instruments were pilot-tested with a sample of 24 Grade 11 mathematics teachers from another district that was not part of the study. The reliability (coefficient alpha values) of the instruments was 0.92 and 0.95 for the self-evaluation and peer evaluation questionnaires respectively. These values implied that the instruments were reliable (Tavakol & Dennick, 2011).

4.3 Data analysis

In this study, teacher and peer responses on the same set of questions were compared. The means and standard deviations of their responses were compared to examine the extent to which teachers' reports of their teaching were consistent with the reports of their peers. The responses of teachers and peers were also correlated to examine the degree of association between the teachers' and the peers' responses; that is to see if teachers' more positive responses to some items corresponded to more positive responses from their peers. A high correlation between the two responses would imply that peer reports of teachers' instructional practice may be used in place of teacher reports in classroom instruction studies.

5. Findings

5.1 How different are teacher self-evaluations and peer evaluations of classroom instruction?

The result of the comparisons of the averages of the teachers' and peers' responses to each of the items is presented in Table 2. The Table shows that on 30 items the means of teacher evaluations were higher than the means of peer evaluations. On the remaining two items (Items 7 and 31), the means of teacher self-evaluation were equal to the means of peer evaluation. The differences between self and peer evaluations means on the 30 items ranged from 0.06 (items 19 and 21) to 0.61 (item 16).

Table 2: Self and peer means evaluations

		Self		Peer	
		Mean	SD	Mean	SD
1	Demonstrated sound knowledge of the curriculum	5.22	0.63	5.00	0.67
2	Demonstrated sound knowledge of students' prior conceptions	5.11	0.74	4.94	0.37
3	Clearly explained procedures for solving problems	5.44	0.50	5.06	0.50
4	Explained concepts in different ways to help students understand	5.11	0.74	4.61	0.57
5	Used a variety of instructional strategies appropriate to the topic	5.00	0.94	4.78	0.82
6	Always related ideas to students' prior knowledge	4.67	0.82	4.50	0.47
7	Planned in advance to ensure lessons run smoothly	4.89	0.74	4.89	0.61
8	Always attended class	5.44	0.68	5.28	0.42
9	Was always punctual for class	5.33	0.67	5.06	0.50
10	Reviewed lessons by connecting to previous classes	5.00	0.82	4.78	0.48
11	Previewed lessons by connecting to future classes	5.22	0.63	5.11	0.57
12	Created opportunity for students' participation in classroom problem solving	5.00	0.94	4.89	0.61
13	Carried the whole class along in lesson delivery.	5.33	0.67	4.89	0.70
14	Asked probing questions when student answer was incomplete	5.22	0.63	4.72	0.63
15	Encouraged students to ask questions	5.56	0.50	5.00	0.62
16	Supported lessons with useful class work	5.78	0.42	5.17	0.75
17	Used different resource materials	5.22	0.63	4.89	0.74
18	Brought lessons home by using examples in the students' context	5.67	0.47	5.44	0.50
19	Built teaching on students' prior knowledge	5.56	0.50	5.50	0.47
20	Gave individual support to students when needed	5.33	0.67	4.89	0.74
21	Adjusted lessons when students experienced difficulties in learning	4.89	0.57	4.83	0.67

22	Challenged students to think through and solve problems, either by themselves or together as a group	5.33	0.47	5.00	0.24
23	Took extra steps to help all students learn and achieve success in maths	5.11	0.74	4.83	0.58
24	Taught in a manner that took cognisance of differing learning abilities of students	5.22	0.79	4.83	0.67
25	Ensured that lessons on the topic were well organised	5.33	0.67	5.11	0.70
26	Provided inspirational learning environment	5.56	0.68	5.00	0.47
27	Well articulated teacher tasks and student tasks	5.22	0.63	4.94	0.68
28	Asked questions to monitor students' progress	5.22	0.79	5.06	0.60
29	Monitored and checked students homework always	5.11	0.57	4.94	0.60
30	Gave feedback to students about their homework and assignment	5.33	0.47	5.11	0.57
31	Appropriately integrated assessment in teaching	5.11	0.57	5.11	0.31
32	Used assessment results to provide extra help to students	5.11	0.31	4.89	0.57

To test whether the difference in self and peer means ratings was significant, a paired t-Test was carried out. The result of the t-Test (Table 3) demonstrates that the difference between the self and the peers' means evaluations is statistically significant ($t=0.000$).

Table 3: Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Self - Peer	.2684375000	.1643385447	.0290512248	.2091871363	.3276878637	9.240	31	.000

5.2 Correlation between teacher self-evaluations and peer evaluations

The results of the correlations between the teacher self-evaluation and peer evaluations of the teachers' teaching are presented in Tables 4 – 8.

5.2.1 Knowledge of the subject content

Table 4 indicates the relationship between the teachers' and their peers' reports of the teachers' knowledge of the subject content. Only the relationship between the teachers' and their peers' report on item K5 ("used a variety of instructional strategies appropriate to the topic") was positive and significant. This implies that the teachers' self-evaluations of their knowledge of the subject content did not agree with their peers' evaluations of the teachers' knowledge of the subject.

Table 4: Correlations between teachers' self-evaluations and peers' evaluations of the teachers' knowledge of the subject content

Variable	r
K1	-.26
K2	-.38
K3	.12
K4	.50
K5	.71*
K6	.56

* Correlation is significant

5.2.2 Lesson presentation

The result in Table 5 demonstrates the relationships between self and peer reports on the teachers' lesson presentation variables. The result shows a high level of discrepancy between the teachers' and their peers' evaluations of the former's lesson presentation as there was no significant relationship between teachers' self-evaluations and the peers' evaluations on any of the variables. This is far from what was expected because the teachers and their peers ought to be very knowledgeable about the subject and lesson presentation and hence were expected to give a concomitant account of the teachers' presentation of their lessons.

Table 5: Correlations between teacher self-evaluations and peer evaluations of teacher lesson preparation

Variable	<i>r</i>
Pr1	.50
Pr2	.14
Pr3	-.38
Pr4	.19
Pr5	.43
Pr6	-.17
Pr7	.47
Pr8	.29
Pr9	.24
Pr10	.00
Pr11	-.21
Pr12	.28
Pr13	.39
Pr14	.34
Pr15	-.31

5.2.3 Lesson preparation

Three variables were used to evaluate the teachers' lesson preparation. The result of correlation analysis between the teachers' self-evaluations and their peers' evaluations based on the variables (Table 6) shows that only on item LP3 ("... was always punctual to class") did the teachers' and peers' evaluations of the teachers' lesson preparation correlate positively and significantly.

Table 6: Correlations between teachers' self-evaluations and their peers' evaluations of the teachers' lesson presentation.

Variable	<i>r</i>
LP1	.15
LP2	.63
LP3	.71*

* Correlation is significant

5.2.4 Lesson Organisation

The correlation between teachers' self-evaluations and their peers' evaluations of the teachers' Lesson Organisation is depicted in Table 7. The Table shows that in no case did the teachers' and peers' evaluations correlate significantly. The finding implies lack of agreement between teachers' and their peers' report of the teachers' lesson organisation.

Table 7. Correlations between teachers' self-evaluations and peers' evaluations of the teachers' lesson organisation

Variable	<i>r</i>
LO1	.11
LO2	.52
LO3	.44

5.2.5 Student assessment

The result of the correlations between self and peer evaluations of the teachers on the student assessment variables (Table 8) indicates that there is no significant relationship between teachers' self-evaluations and their peers' evaluations on any of the student assessment variables. This implies general disagreement between teachers' and their peers' reports on the teachers assessment of the students.

Table 8: Correlations between self and peer reports on assessment of students

Variable	r
SA1	.08
SA2	-.22
SA3	-.13
SA4	-.06
SA5	-.55

6. Discussion and Conclusion

The findings of the study establish that there is a significant difference between the teachers' self-evaluations of their instructional practices and the peers' evaluations. The teachers rated their instructional practices higher than their peers rated them. Further analysis demonstrates that the teachers' self-evaluations of their instruction and their peers' evaluation of the teachers' instruction generally did not correlate significantly. One possible reason for the findings is that the teachers did not give true evaluations of their teaching because they held exaggerated views of it. This may be the reason why, in almost all the items, the teachers' means evaluations of their instruction practices were higher than the peers' means evaluation. It is also possible that teachers wanted to impress the researcher by rating their teachings very positively so as to appear socially desirable (Mayer, 1999).

The discrepancy between the teachers' self-evaluations and the peers' evaluations could also exist because some of the peers gave malicious evaluations of their colleagues' teaching, using the evaluations as an opportunity to victimise them. Similarly, it could be the case that the peers, being experts in the subject matter, had very high expectations of the teachers' instructional quality and hence were hardly impressed by the teaching instruction practices. These reasons may explain why the peers' mean evaluations were less than the teachers' mean evaluations on 30 items out of the 32.

Moreover, despite being told that the study was for academic purposes and was not meant to be used for their promotion by the Department of Basic Education or any other administrative issue, some of the teachers might have thought that their deficiencies would be exposed to the Department and hence decided to provide false evaluations of their teaching. On the other hand, it is possible that the peers under-evaluated the teaching because the teachers' instructional practices did not conform to the peers' style of teaching.

These findings have numerous implications for the use of teachers' self-evaluation of their teaching and peer evaluation of this teaching in educational research and in taking policy decisions. Often, researchers are constrained to use surveys as the easiest means of data collection; this urge has to be weighed against the potential pitfalls of the use of survey data in studying teaching and also against the significance of the study.

Generally, self and peer evaluations of teaching practices can afford an indication of what transpired in the classroom and hence provide valuable insights into teachers' work, offering expert guidance about how teaching can be improved. However, survey studies may not furnish accurate accounts of educational practices on which high-stakes decisions can be taken. To ensure credibility, high-stakes studies should not depend exclusively on survey data but on a triangulation of different data sources and data collection methods. The triangulation of multiple-methods and multiple-sources as supported by teacher evaluation experts (Ory & Ryan, 2001) will most likely eliminate the potential bias posed by each of the methods and sources and constitute a more objective measure of instructional practices and quality.

References

- Berk, R. A. (1979). The construction of rating instruments for faculty evaluation. *Journal of Higher Education*, 50: 650-669.
- Berk, R. A. (2005). Survey of 12 Strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62.
- Chen, Y. & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71-88.
- Cranton, P. (2001). Interpretive and critical evaluation. In C. Knapper & P. Cranton (Eds.), *Fresh approaches to the evaluation of teaching: new directions for teaching and learning*. San Francisco: Jossey-Bass.
- Denson, N., Loveday, T. & Dalton, H. (2010). Student evaluation of courses: what predicts satisfaction? *Higher Education Research & Development*, 29(4), 339-356.
- Desimone, L. M., Smith, T. M. & Frisvold, D. E. (2010). Survey measures of classroom instruction comparing student and teacher reports. *Educational Policy*, 24 (2), 267-329.
- French-Lazovik, G. (1981). Peer review: documentary evidence in the evaluation of teaching. In Jason Millman (Ed), *Handbook of Teacher Evaluation*. London: Sage.

- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R. & Stigler, J. W. (2012). Measuring usable knowledge: teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568-589.
- Ko, J., Sammons, P. & Bakkum, L. (2013). *Effective teaching: a review of research and evidence*. Berkshire: CfBT Education Trust.
- Lampert, M. (2010). Learning teaching in, from, and for practice: what do we mean? *Journal of Teacher Education*, 61(1-2), 21-34.
- Marsh, H. W. & Hocevar, D. (1991). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching & Teacher Education*, 7: 9-18.
- Mayer, D. P. (1999). Measuring instructional practice: can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29-45.
- Ogbonnaya, U. I. & Mogari, L. D. (2011). Development and validation of student evaluation instrument for measuring effective mathematics teaching. Proceedings of 19th Annual Conference of the Southern African Association for Research in Mathematics, Science and Technology Education, University of North West Mafikeng, 225 – 233
- Ory, J. C., & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In T. D. Johnson & D. L. Sorenson (Eds.), *Student ratings of instruction: new directions for teaching and learning*. San Francisco: Jossey-Bass.
- Patton, M.Q. (1987). *Qualitative research evaluation methods*. Thousand Oaks, CA: Sage Publishers.
- Praetorius, A., Pauli, C., Reusser, K. & Rakoczy, K. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2-12
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- Trochim, W. M. K. (2006). Research knowledge base. Assessed February 18, 2011 from <http://www.socialresearchmethodsnet/kb/scallik.php>.